DOCUMENT RESUME

ED 276 711 SP 028 358

AUTHOR	Ashton, Patricia; And Others
TITLE	Does Teacher Education Make a Difference? A Literature Review and Planning Study. Executive Summary and Technical Monograph.
SPONS AGENCY	Florida State Dept. of Education, Tallahassee. Student Assessment Section.
PUB DATE	86
GRANT	050-94640-850000
NOTE	270p.
PUB TYPE	Information Analyses (070)
EDRS PRICE	MF01/PC11 Plus Postage.
DESCRIPTORS	*Academic Achievement; Higher Education; Preservice Teacher Education; *Program Effectiveness; Teacher Certification; *Teacher Education Programs; Teacher Evaluation

ABSTRACT

Research was reviewed that addressed the question: Is type of teacher education related to student performance? Major findings were: (1) teachers with master's degrees were rated as more effective by supervisors and had higher levels of student achievement than teachers with bachelor's degrees; (2) supervisors rated college of education graduates more highly than graduates from liberal arts; (3) teachers who earned more credit hours in professional education obtained higher ratings from supervisors and had higher student test scores than teachers with fewer credits; (4) number of credit hours taken by teachers in academic subjects was reflected in their students' achievement; (5) teachers with higher grade point averages and higher scores on tests in the subjects they taught had higher student achievement; (6) the National Teacher Examination was not a good predictor of either teacher performance or student achievement; (7) teachers' grade-point average tended to be a more stable predictor of teacher performance than teachers' scores on a single test; and (8) teachers meeting certification requirements received higher supervisor ratings and had higher student achievement than teachers who did not meet certification standards. Methodological weaknesses in the studies were identified, and a design for future research using causal modelling was proposed. A 12-page reference list and tables summarizing the research studies under various headings are appended. (Author/AA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED276711

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

DOES TEACHER EDUCATION MAKE A DIFFERENCE?

A Literature Review and Planning Study

by

Patricia Ashton Linda Crocker
University of Florida

and

Stephen Olejnik
University of Georgia

Executive Summary and Technical Monograph

Prepared for the Institute on Student Assessment and
Evaluation for Grant No. 050-94640-850000,
Florida Department of Education, 1986.

FOREWORD

This study was undertaken at the express request of Robert Graham, Governor of Florida, to provide information to state-level policy-makers concerned with improving the quality of teaching and teacher-education. The Institute for Student Assessment and Evaluation at the University of Florida provided financial support for this project. The Institute is supported by a grant from the Florida Department of Education. The views expressed here are the authors' and do not necessarily reflect those of the funding agencies.

We wish to acknowledge the ready assistance of Debra Gallagher and Tsao Mei-Jung in locating the many documents reviewed for this project. Julie Morin and Nancy Klich helped to compile the reference list and appendix tables. Major credit for production of the final version of the manuscript goes to Elizabeth Ann Franks who mastered the intricacies of the word processor with unparalleled skill and whose organizational and editorial skills brought the project to conclusion on schedule.

Finally we are indebted to the following panel of distinguished professional colleagues who reviewed the manuscript and offered invaluable constructive suggestions:

James Algina
Professor and Chair
Foundations of Education
University of Florida

Kern S. Alexander
President
Western Kentucky University

Eileen Castle
Principal
Polk County Public Schools

Natafly S. Glasman, Dean
Graduate School of Education
University of California
Santa Barbara

Suzanne Kinzer
Planner
Broward County Public Schools

Sue M. Legg
Associate Director
Office of Instructional
Resources
University of Florida

Mary Rohrkemper
Assistant Professor
Bryn Mawr College

Jeaninne Webb
Director
Office of Instructional
Resources
University of Florida

EXECUTIVE SUMMARY

DESCRIPTION OF THE STUDY

The purpose of this project was to review the research literature and conduct a planning study addressing the question: Is type of teacher training related to student performance? These aspects of teacher training were considered in the literature review: level of degree (e.g., bachelors or masters), field in which degree was obtained, and teacher certification status. If the review indicated a lack of definitive information in these areas, a design for future research was also to be included.

The basic methodology of this project was to survey professional literature in education and social sciences that described research on the relationship between teachers' formal education and their competence in professional practice. In all, over 200 articles, books, dissertations, and research reports related to this topic were located and reviewed. More than 135 of these resources were selected for citation in the report. Further information on recent efforts in this area was obtained by conference attendance and consultation with prominent educational researchers and policy-makers. The organization of the resulting literature review is shown in Chart 1.

CHART 1

I. Introduction and history of major types of teacher training programs in the U. S.

II. Are teachers with master's degrees more effective?

III. Does professional education make a difference?

    1. Comparisons of liberal arts and education graduates

    2. Effects of coursework in professional education

    3. Effects of coursework in academic subject areas

IV. Does teachers' demonstrated knowledge of a subject affect their performance?

V. Does teacher certification make a difference?

VI. Methodological issues

VII. A proposed study

GENERAL FEATURES OF STUDIES REVIEWED

There was great variation in method and design among the wide range of studies reviewed, but several patterns occurred across a number of studies. One common approach was to obtain student achievement test scores and to determine which teacher, school, and student characteristics found in school records could be used to predict those scores using multiple regression analysis. Another approach was to ask school principals or superintendents to identify outstanding and unsatisfactory teachers and then identify characteristics that distinguished between these two groups.

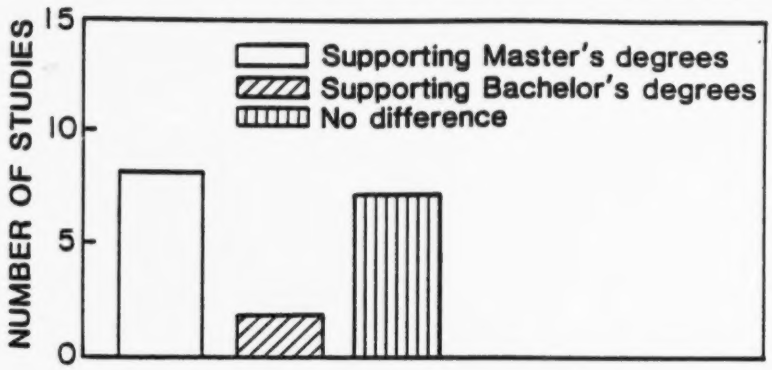
A few researchers identified groups a priori on the characteristic of interest (e.g., certification status) and then systematically collected follow-up data on teacher performance in the classroom or their students' achievement. The number of teachers studied ranged from as few as 18 to as many as 1200. Typically fewer than 100 teachers were included in any single study.

MAJOR CONCLUSIONS OF THE LITERATURE REVIEW

- Teachers with master's degrees are rated as more effective by their supervisors and have higher levels of student achievement. Based on the fairly stringent statistical criteria used to declare that a finding is significant, only 1 out of 20 studies is expected to show a positive relationship due to chance. Chart 2 shows that 8 out of 15 studies showed a significant positive relationship between level of educational degree and teachers' classroom performance or their students' achievement. Moreover, the 4 studies that were strongest in terms of research quality all showed positive relationships between level of teacher education and teaching effectiveness.

Chart 2.

Results of Studies of Level of Education



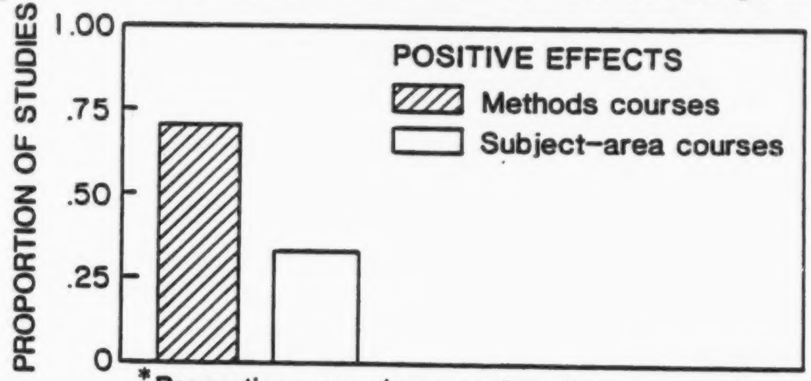
- Graduates of colleges of education are more highly rated by their supervisors than graduates from liberal arts or other non-education majors. Principals' ratings of education and liberal arts majors were compared in 3 studies, and in each study the education majors received higher ratings than the liberal arts majors.

- Teachers who earn more credit hours in professional education obtain higher ratings from supervisors and have higher student test scores than teachers with fewer credits in professional education. Chart 3 shows that 5 out of 7 studies demonstrated a positive relationship between the amount of professional coursework and teacher effectiveness criteria. This suggests that when teachers receive instruction in how to teach a subject, it has a positive impact on their teaching effectiveness.

- There is weaker evidence that the number of credit hours taken by teachers in academic subjects is reflected in their students' achievement. Only 5 of 16 studies showed a positive relationship between the number of credits teachers earn in academic fields and their teaching effectiveness. The majority of studies failed to support the hypothesis that increasing teachers' subject-area preparation will improve their students' performance.

Chart 3.

Results of Studies of Credits Earned in Teaching Methods and Academic Subjects



*Proportions are shown rather than raw numbers because they are based on different numbers of studies.

- Teachers with higher grade point averages and higher scores on tests in the subjects that they teach tend to have higher student achievement, especially among high- achieving students and on tests of higher-order thinking skills. The relationship between teachers' scores on subject area tests and their students' achievement was investigated in 14 studies. In 9 of those studies, there was evidence of a positive relationship. The relationship between teachers' GPA and their teaching effectiveness was examined in 5 studies, and each time a small but significant positive relationship was reported.

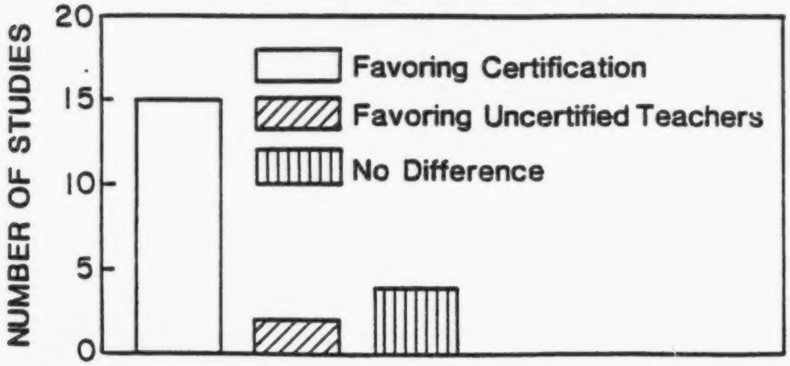
- The National Teacher Examination is not a good predictor of either teacher performance or student achievement. Of 14 studies that examined the relationship between teachers' NTE scores and teacher effectiveness, only 5 showed any evidence of a relationship.

- Teachers' grade-point average tends to be a more stable predictor of teacher performance than teachers' scores on a single test. Fourteen studies were examined that used subject matter tests as the criterion of teacher effectiveness. Of these studies, 9 showed evidence of a relationship between teacher knowledge and teacher effectiveness. In contrast, each of the 5 studies using teachers' GPA as the criterion of teacher knowledge showed evidence of a significant relationship.

- Teachers meeting regular state certification requirements consistently receive higher supervisor ratings and have higher student achievement than teachers who do not meet certification standards. Chart 4 shows the results of the studies that have examined the relationship between teacher certification and teacher effectiveness. Fourteen of the 19 studies favored teachers holding regular certification; only 2 studies favored uncertified teachers, and in 4 of the 19 studies, no differences were found between certified and uncertified teachers. Certified teachers also remain in teaching as a career longer than uncertified teachers.

Chart 4.

Results of Certification Studies



MAJOR WEAKNESSES OF PREVIOUS RESEARCH

Much of the research conducted to date has been fraught with methodological weaknesses. The prevalence of these weaknesses among the studies reviewed limits the confidence that can be placed in these findings when drawing implications for policy or practice. The weaknesses noted in the existing body of research on effects of teacher preparation stem from three sources: (1) researchers used conveniently available data rather than collecting data in the form needed; (2) recently developed statistical procedures needed for appropriate data analyses were not widely available, when many of these studies were conducted; and (3) the scope of the study and sample were restricted because of inadequate resources. Some common problems have been

- ✓ Sampling bias occurred in selection of teachers or inadequate numbers of teachers or schools were sampled to permit detection of effects at the classroom level.
- ✓ Teacher educational data were not collected or reported in sufficient detail to permit inferences that could guide future policies on teacher education.
- ✓ Control for prior level of student achievement was inadequate.
- ✓ Inadequate experimental or statistical controls for the effects of intervening variables (e.g., student SES, school characteristics, and teachers' level of motivation or sense of efficacy) that exert major influence on student achievement were incorporated into the studies.
- ✓ Studies have been limited in scope, focusing only on one outcome measure or one grade level. Student attitudes have seldom been considered.
- ✓ Student performance within a single study has been measured with different tests so that equating these scores is questionable.
- ✓ Principal ratings (which are highly subjective) have often served as the outcome variable rather than objective measures of student outcomes.
- ✓ Data were inappropriately analyzed using student score or school average as the unit of analysis. The most appropriate level of analysis, however, is the class average when inferences are to be drawn about effects of teacher characteristics.

- √ The effects of correlated variables such as teacher ability, experience, teacher level of education, and teachers' salary frequently were confounded and the method of statistical analysis employed did not permit separate estimation of the effects of these variables.

- √ Previous input-output studies of educational effects included too many variables in regression analysis. Such a shotgun approach cannot significantly improve our understanding of how teacher education influences teachers' classroom effectiveness.

Conclusion

There is a clear need for a large-scale, comprehensive study of the relationship between critical variables in teacher preparation, school characteristics, and student performance. This research should take advantage of current state-of-the-art methodology for addressing this question.

THE PROPOSED STUDY

In light of this review, the future studies of the impact of teacher education on student achievement should consider:

1. Teacher personal characteristics (i.e., social class, race, and verbal ability);

2. Teacher job-related characteristics (i.e., experience in teaching, and sense of efficacy);

3. School characteristics (i.e., principal's level of education and institutional leadership, per-pupil expenditure on insturctional materials, size, teacher turnover);

4. Class ascribed characteristics (i.e., race, socioeconomic status);

5. Class school-related characteristics (i.e., prior achievement).

Furthermore these variables should be considered within the framework of a sound theoretical model which provides a coherent approach to data collection and analysis. This model should provide for methods of testing both the direct and indirect effect of these variables on student performance while controlling for effects of other variables in the model. Until recently statistical procedures for accomplishing this were not generally available to educational researchers. Figure 1 presents a diagram that illustrates how teacher demographic characteristics, teacher education characteristics, school factors, and student characteristics combine to influence student achievement. Such a diagram is the first step to development of a model of that can be used to assess the impact of teacher education on student achievement.

Research Questions to be Addressed

Several variations on the model shown in Figure 1 could be tested to find which seems to offer the best fit to actual teacher/student data. In addition to testing the overall fit of the data to the model, a series of questions following the paths depicted by the arrows in Figure 1 would be answered. For example, one series of questions would be:

> To what extent does teacher social class directly influence the level of education attained by the teacher?
>
> What is the direct effect of teachers' level of education on teachers' sense of efficacy?
>
> What is the direct effect of teachers' sense of efficacy on student achievement?
>
> What are the direct and indirect effects of teachers' level of education on student achievement?

Additional sets of questions would be answered for each possible path shown by the arrows that connect the variables in the model.

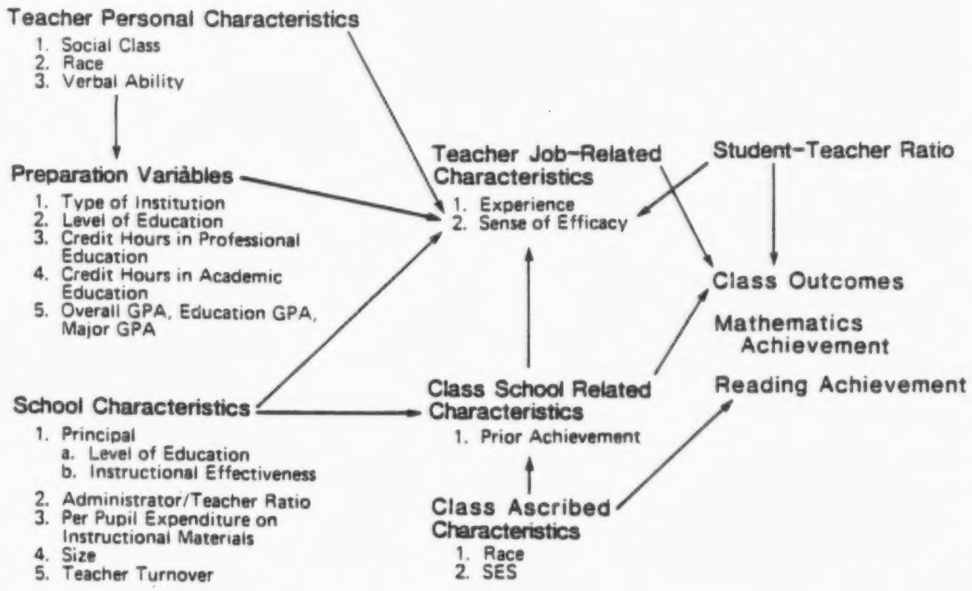

Teacher Personal Characteristics
1. Social Class
2. Race
3. Verbal Ability
Preparation Variables
1. Type of Institution
2. Level of Education
3. Credit Hours in Professional Education
4. Credit Hours in Academic Education
5. Overall GPA, Education GPA, Major GPA
School Characteristics
1. Principal
a. Level of Education
b. Instructional Effectiveness
2. Administrator/Teacher Ratio
3. Per Pupil Expenditure on Instructional Materials
4. Size
5. Teacher Turnover
Teacher Job-Related Characteristics
1. Experience
2. Sense of Efficacy
Student-Teacher Ratio
Class Outcomes
Mathematics Achievement
Reading Achievement
Class School Related Characteristics
1. Prior Achievement
Class Ascribed Characteristics
1. Race
2. SES

Figure 1.

STRUCTURAL MODEL OF TEACHER PREPARATION VARIABLES INFLUENCING CLASSROOM OUTCOMES

Instruments and Methods of Data Collection

As part of the present study we explored the feasibility of collecting accurate and timely information on teacher education and student achievement variables from data currently available through the State Department of Education. Data available from the Teacher Certification Office received particular scrutiny. We also consulted with school district personnel in several regions of the state to identify pragmatic procedures useful for collecting student achievement and teacher educational data at the district level. This information was taken into account in formulating the proposal.

From the feasibility study, we determined that the Comprehensive Test of Basic Skills (CTBS) is the most widely used standardized achievement test in the school districts in Florida. Student achievement test data can be collected from the county district office in the form of individual student test scores or the average test score for a given classroom. We also learned that the detailed information needed on teacher educational background cannot be obtained from existing data files of the Department of Education. Accurate, complete teacher educational background data can best be obtained from teachers directly. Furthermore, there is considerable variance in educational preparation of Florida's elementary teachers, but not in certification status of elementary teachers in Florida (Scott & Damico, 1985), so it is most reasonable to concentrate on differences in the type and amount of teachers' educational preparation.

The following instruments or methods of data collection would be employed:

1. A standardized achievement test with subscores in math and reading, such as the Comprehensive Test of Basic Skills, presently administered in 32 Florida counties;

2. A teacher questionnaire containing items relevant to the teacher's demographic and educational background;

3. A standardized measure of teachers' sense of efficacy (or motivation) such as that developed by Gibson & Dembo (1984). Scores on this instrument are a function of teachers' confidence in their own ability to teach and students' abilities to learn and are related to teacher and student behavior and student achievement (Ashton & Webb, 1986).

4. A school questionnaire to be completed by the principal containing items on the school-level variables and classroom-level variables.

All items on these questionnaires would be pilot-tested for clarity of meaning and ease of response in at least two schools before being used in the field study. Teacher and school questionnaires would be distributed and collected by the research team on site in the schools.

Sample

Appropximately 200 teachers at second and fifth grades are needed for the investigation. The 200 teachers at each grade level would be selected using a stratified sampling procedure so that approximately one-third should be from rural school districts, one-third from small metropolitan, and one-third from large metropolitan communities. A multistage sampling plan would be followed to select districts.

The minimum sample size was determined on the basis of several factors including a) the minimum effect size to be detected which would be judged important from a practical point of view; b) the number of independent variables under investigation; c) the desired power level and d) the criterion for statistical significance.

While the choice of grade levels is arbitrary the selection of an early elementary and a late elementary grade level is recommended. The rationale for the choice is that (1) elementary grade level instruction is based on intact classrooms, (2) previous research has focused at these levels of instruction, (3) the grade level spread provides an opportunity to explore the generalizability of the results, (4) the inclusion of an early elementary grade reduced the confounding of multiple teacher effects; and (5) the inclusion of the upper elementary grade will increase the variability in achievement test scores across classrooms.

Data Analysis

The data analyses would include calculations to describe the data distribution in terms of means and standard deviations for continuous variables and proportion of response frequencies for categorical variables.

Further analyses would be conducted using LISREL VI, a program authored by Joreskog and Sorbom (1985), for the analysis of linear structural relationships. Specifically, the analysis would be used to (1) determine whether there is adequate fit between the data and the model(s) and (2) test the significance of coefficients which quantify the relationship between the outcome variables (e.g., student achievement) and other variables in the model. The strength of this procedure lies in its ability to yield quantitative estimates of the direct and indirect relationships student achievement and teacher preparation while taking into account the complex relationships between other variables in the model. Another strength of the study is that it would be replicated at two grade levels and in two subject areas (math and reading).

Second Phase of Research

In the event that promising relationships are revealed we recommend that a second phase of research explore the causal nature of the relationships through experimental research. For example, a limited number of randomly selected master's degree teachers could be compared with bachelor's degree teachers as they instruct their students in similar school settings on one or more common unit(s) of material prepared or selected specifically for this study. The design should permit in-depth observations of teacher and student behavior as well as student achievement and student attitudes. Assessment of achievement would include both lower and higher order cognitive skills. Further, we recommend that the stability of the effects be examined across grade level and subject matter.

Time and Cost Estimates

The total time required to conduct a project such as that described, would be approximately 18 months. During the first 12-month period it would be reasonable to accomplish the major tasks of organizational start-up, questionnaire development, pilot-testing, drawing the sample, securing cooperation of participating districts and schools, collecting the teacher data, and obtaining student test-score data. The next 6 months would be devoted to data analysis and preparation of the final report.

The estimated cost for supporting the activities of the first 12 months would be approximately $85,000. The cost for supporting the major activities of the last 6 months of the 18-month project would be an additional $25,000. Thus the total cost of the 18-month project would be approximately $110,000. These cost estimates are based on the assumption that the work be conducted at one of the state universities or by an organization that would not charge for indirect costs.

DOES TEACHER EDUCATION MAKE A DIFFERENCE?

A Literature Review and Planning Study

Technical Monograph

Prepared for the Institute on Student Assessment and
Evaluation for Grant No. 050-94640-850000,
Florida Department of Education, 1986.

CONTENTS

1. INTRODUCTION 1

Defining Effective Teaching 2
History of Teacher Preparation in the U.S. 3

2. ARE TEACHERS WITH MASTER'S DEGREES MORE EFFECTIVE? . 9

Effects on Individual Student Performance 10
Studies Using Teacher (Classroom) as the Unit of Analysis 19
Studies Using School/District as the Unit of Analysis 23
Summary and Implications 39

3. DOES PROFESSIONAL EDUCATION MAKE A DIFFERENCE? . . . 43

Characteristics of Academic Institutions and Teacher Effectiveness 43
Amount of Coursework in Professional Education . . . 51
Coursework in Academic Subject Areas 58
Summary and Implications 67

4. DOES TEACHERS' KNOWLEDGE OF SUBJECT AFFECT THEIR PERFORMANCE? 70

Studies Using the National Teacher Examination . . . 70
Studies Using Other Tests of Teacher Knowledge . . . 74
Using Grade-Point-Average as a Measure of Teacher Knowledge 84
Summary and Implications 87

5. DOES TEACHER CERTIFICATION MAKE A DIFFERENCE? 90

Studies under Naturalistic Classroom Conditions. . . 92
Studies under Controlled Classroom Conditions. . . . 100
Studies using Certification Measures 104
Summary and Implications 109

6. METHODOLOGICAL ISSUE 112

Sample Selection 113
Inadequate Specification of Teacher Education Data. 114
Controlling Extraneous Variables 115
    Student's Family Background 115
    Student Characteristics 116
    School-related Student Characteristics 117
    School Inputs 118
Limitations of Cognitive Measures 121
Ratings as a Criterion of Teacher Effectiveness . . 123
Stability of Teacher Effectiveness 124
Unit of Analysis 130
Linear Analyses and Interaction Effects 133
The Shotgun Approach to Data Analysis 134
A Methodological Synthesis 135
        The Conservatism of Input-Output Research 136
        The Inability to Draw Causal Inferences from Input-Output Data 136
        The Power of Combining Experimental and Nonexperimental Designs 138
Summary and Implications 140

7. A PROPOSED STUDY 143

The Model 144
The Questions 147
Instruments and Methods of Data Collection 148
Sample 150
Data Analysis 154
Second Phase of the Research 155
Time and Cost Estimates 155

8. REFERENCES 157

9. APPENDIX 169

CHAPTER 1

INTRODUCTION

The purpose of this paper is to address the global question: Does teacher-education make a difference? A review of the research literature was undertaken to address the following three components of this broad question that have direct implication for educational practice and policy decisions:

1. Is there any evidence that teachers with master's degrees are more effective than teachers with baccalaureate degrees?

2. Is there any evidence that formal training in pedagogy (i.e., methods of teaching commonly offered in colleges of education) produces more effective teachers than a liberal arts education?

   A. Are graduates of colleges of education more effective teachers than graduates of liberal arts colleges?

   B. Is there a relationship between number of college credits earned in professional education courses and teacher effectiveness?

   C. Is there a relationship between number of college credits earned in the subject area and teacher effectiveness in teaching that subject?

3. Is there any evidence that teacher knowledge in a subject (as measured by test scores or academic grade-point average) is related to teacher effectiveness?

4. Is there any evidence that certified teachers are more effective than teachers who are not certified in their respective fields of instruction?

From the onset it should be apparent that these simply-phrased questions represent gross oversimplification of complex issues. It seemed unlikely that the present review would locate studies that could provide definitive answers when considered separately. Because of the importance of the questions, however, our purposes were to (a) review those empirical studies that have bearing on these issues; (b) critically evaluate these studies so that their results might be interpreted with appropriate caution; (c) summarize findings across multiple studies to obtain a clearer picture of recurrent results that can be interpreted with some confidence; and (d) identify needs for additional research in this area.

Defining Effective Teaching

One reason that teaching has so many critics is simply that it is the one profession with which almost everyone has some familiarity. From kindergarten through the twelfth grade, a typical citizen in our society has the opportunity to observe from 20-30 members of this profession in daily practice 6 hours daily for 180 days per year. Thus from personal experience nearly everyone who has ever attended school has formed some impression of effective teaching. Consequently, there are no universally accepted definitions of effective teaching among laymen or within the profession itself. In a recent review of literature on teacher evaluation, Darling-Hammond, Wise, and Pease (1983) differentiated between teacher competence, teacher performance, and teacher effectiveness as follows:

> 1. Teacher competence refers to the knowledge and skills a teacher possesses;
>
> 2. Teacher performance refers to actual teacher behaviors in the classroom (i.e., what the teacher does on the job);
>
> 3. Teacher effectiveness refers to the effects of teacher performance on students.

This distinction between competence, performance, and effectiveness seemed a useful one to make in the present review.

While it is natural to regard teacher effectiveness, in terms of student achievement, as the ultimate criterion in evaluation of teacher preparation programs, a number of authors and researchers have pointed out the difficulties in attempting to establish such a relationship. In reviewing studies of the effects of teacher preparation, it is important to recognize that some researchers have elected to study the effects of teachers' academic preparation on competence, while others have chosen performance, and still others, effectiveness (as defined by student achievement test performance) as their outcome variable. To summarize results of these studies without making such distinctions would invite confusion and misinterpretation of their findings. In this review, we have focused primarily on studies in which student achievement was used as the ultimate criterion of teacher effectiveness; in cases where more immediate or intermediate criteria were used (i.e., teacher knowledge or classroom behavior), this has been carefully noted.

History of Teacher Preparation in the U.S.

Before tackling questions about the comparative effectiveness of various types of teacher preparation, it is helpful to have some historical perspective of how current professional educational programs developed. Cubberly (1919) noted that prior to the mid-nineteenth century the major qualification for a teacher was "soundness in faith." No other qualities were considered as important, although some modicum of literacy presumably was expected. Class (1931) supplied the following picture of the development of teacher preparation programs beginning in the early 1800s. The first school established expressly for the purpose of providing professional training for teachers beyond rudimentary elementary school education was opened in Concord, Vermont in 1823 by the Reverend Samuel Hall. This institution, called an academy, offered education on a par with that offered by secondary schools of that era. The typical academy curriculum included offerings in areas such as English, mathematics, history, navigation, theology, sciences, political economics, and the art of teaching. Samuel Hall's course on the art of teaching was based on his monograph entitled "Lectures on School Keeping" (which was widely used in its time) and upon demonstration of teaching methods using a class of children maintained at the academy for that purpose. By 1830 Hall's academy had moved to Andover, Massachusetts and had been copied by institutions in a number of communities, particularly in New England. In most instances, an academy offered a three-year program beyond elementary school.

During this same period the first public high school opened in Boston in 1821. The public high school curriculum was quite similar to that of the academies described above with the exception that Latin and Greek were mainstays of this curriculum while courses on the art of teaching were generally lacking. Although graduation from the three-year public high school or an academy was regarded as more than sufficient preparation for the education of teachers in many communities, at the high school level these teachers were barely more literate than their own students.

Recognizing this, in 1825, Thomas Gallaudet proposed the need for an institution of post-secondary education for the training of classroom instructors just as there were institutions dedicated to the professional preparation of students of divinity, law, and medicine at that time. Horace Mann, Secretary of the Board of Education in Massachusetts, was a convert to this viewpoint and under his leadership three institutions for the preparation of teachers were founded in Massachusetts, beginning in 1830. These teacher-education institutions were known as normal schools. Typically the normal school curriculum included studies in reading, grammar, logic, arithmetic, history, geography, physiology, natural sciences, and principles of "ethics and morality" as well as courses in theory and history of education, methods of instruction, school law, and school organization. By 1865, most normal school programs required two years for completion. A substantial focus of the subject matter courses consisted of review of basic materials which the students would be expected to teach and an opportunity to complete exercises in teaching in the experimental or model schools which were maintained by the normal school to provide prospective teachers with some opportunities for observation and classroom experience.

By the early twentieth century, normal schools were being supplanted by teachers colleges as the major avenue for preparation of public school teachers. The first teachers college was opened in 1903 in Ypsilanti, Michigan soon followed by establishment of a number of similar institutions particularly concentrated in the midwest and southern states. Teachers' colleges were distinct from other forms of educational preparation for teachers in that they required graduation from high school for admission (or demonstration of an equivalent level of competence) and offered a four-year course of study leading to a baccalaureate degree. (Presumably, studies in various subject areas were at a more advanced level than those of normal schools.) It is important to note, however, that teachers' colleges of this era also offered two-year and three-year programs of study. Typically prospective elementary school teachers might opt for the two or three year program with the four-year course primarily pursued by intending secondary school teachers. The rapid growth in acceptance and demand for teachers' college programs can be seen from the following statistics: In 1919-1920, there were 46 teachers' colleges and 137 normal schools in the U.S. By 1927-28, the number of teachers' colleges had increased to 137 while the number of normal schools had declined to 69.

In Florida, the development of teacher education programs generally paralleled the national scene. Keck (1985) has chronicled historic events in Florida teacher education in detail, and some of the highlights of her presentation are presented here. In 1851, the Florida legislature authorized creation of two seminaries of learning to provide formal training to both male and female students desiring to become classroom teachers. The East Florida Seminary was established in Ocala one year later. It later moved to Gainesville in 1861. West Florida

Seminary was founded in Tallahassee, but not until five years later. In 1861, the West Seminary was conferred as a military and collegiate institution. The provision of education for black teachers followed in 1866 with the creation of an institution later to become Edwin Waters College, at Jacksonville.

East Florida Seminary offered a typical three-year normal school curriculum in the 1880's and 1890's. In 1905, teacher education became a formal major offered at the University of Florida in the School of Pedagogy. Later this program was included in the College of Arts and Sciences. In 1912, the responsibility for education of teachers shifted to the Teacher's College and Normal School. The normal school program was discontinued in 1928 and in 1931, the present-day College of Education was established under the leadership of Dean James Norman.

The teacher-training program at Tallahassee followed a similar pattern. In 1905, the School for Teachers offered one of four main programs at the Florida State College for Women. The first director was L. W. Buchholz. From 1912-1916, this program was housed in the College of Arts and Sciences, completing the transition to a separate College of Education in 1928.

One interesting aspect of Florida teacher training was that in 1915, the legislature passed a law to support the education of teachers in public high schools. This practice was discontinued in 1931, at the time that Colleges of Education received full-fledged status at the two major state universities.

In 1936, the state of California instigated a new national trend by requiring all teachers in public schools to have four years of academic preparation beyond high school. New York, New Jersey, and Arizona soon required three years of post-secondary education. Clearly the days of the normal school were numbered as an accepted institution for the preparation of teachers.

Within the last 50 years, the baccalaureate degree has emerged as the generally-accepted minimum educational qualification for entry into teaching. Teachers colleges have been absorbed or evolved into colleges or universities with broader curricula offerings. Moreover, increasing proportions of classroom teachers hold graduate degrees at the master's or even doctoral level, and district salary schedules typically award additional pay for attainment of these higher levels of education. Thus it seems quite appropriate for public policy-makers to inquire how much students in public schools benefit from the practice of teachers' pursuit of graduate education. Equally appropriate are questions concerning the amount (or balance) of training in subject areas and method-oriented, pedagogical courses. Chapters 2, 3, and 4 of this monograph provide an overview of research literature concerning these issues. The fifth chapter focuses on the impact of teacher certification requirements in terms of student educational benefits.

CHAPTER 2

ARE TEACHERS WITH MASTER'S DEGREES MORE EFFECTIVE?

Surprisingly few studies have focused primarily on the relationship between the level of education attained by teachers and their students' classroom attainment. It is more common to find that teacher's educational level has been included as one of many teacher variables in a study that focused on other questions (e.g., equality of educational opportunity, Coleman et al., 1966). In some of these cases, however, when level of education was found to be unrelated to teacher effectiveness, this aspect of the study may not have been described adequately to allow critical evaluation or interpretation of the reported finding. In spite of this, there remain a number of studies in which the relationship between teachers' level of education and their teaching effectiveness was examined and described in sufficient detail to be eligible for inclusion in this review.

To synthesize findings from these separate studies it was important to recognize that different researchers used different units of analysis in collecting and analyzing their data. Some have used individual students, some have used classroom means, and some have used school or district-level means. Knowing the unit of analysis is critical to interpretation of research findings because using a different unit of analysis results in addressing a different research question and introduces the opportunity for different methodological problems to occur. Thus studies using different units of analysis have been considered separately in the review that follows. Each section includes a description of the question asked when a specific unit of analysis is chosen, illustrations of policy implication(s) related to the question, an overview of the common methodological problems which may affect interpretation of study findings and summaries of the studies themselves.

Effects on Individual Student Performance

If the data are analyzed using individual student as the unit of analysis, then the question is: Is the achievement of an individual student typically higher when instructed by a teacher with a master's degree than when instructed by a teacher with a bachelor's degree? This question would seem to have enormous practical significance for students, parents, teachers, administrators, and policy makers. After all, if it could be demonstrated that on the average, student performance is greater when students are enrolled in the classroom of a teacher with a master's degree, every parent would want his or her child taught by teachers with advanced degrees. Unfortunately, the inferential statistical methods commonly available to test the differences between the average performance of children taught by bachelor's and master's degree teachers require certain assumptions that are almost inevitably violated in the design of these comparison studies. Specifically, it is the assumption of independence of the observations (i.e., student achievement scores) that is violated. Strictly speaking this assumption could be met only if, for each teacher in the study, the researcher could randomly select one and only one child from that teacher's class. Thus the sample would consist of children who were each taught by a different teacher. In all the studies we reviewed, in which student was the unit of analysis, this assumption was violated because the researchers included all of the children in each teacher's classroom to achieve an adequate sample size. While the approach tends to yield overly liberal results, the extent of its effect on the outcome of any given data analysis can not be fully determined. Furthermore, the problems of analysis and interpretation are compounded when values for some student variables (e.g., SES or prior level of achievement) are individually entered in the analysis for each student, and teacher- or school-level variables have common values for groups of students within the sample. Burstein (1980), Cooley, Bond, and Mao (1981), and Goldstein (1985) are among the many researchers who recently have pointed out misinterpretations and problems that can arise from such analyses. For this reason it is generally more appropriate for policy-based studies of a classroom variable to employ an aggregated unit of analysis, such as teacher or school/district, rather than the student. (Research of this type is reviewed in the next sections.) Thus we point out that all of the following studies to be reviewed in this section suffer from this methodological flaw. In general these studies are discussed below in chronological order of their occurrence.

More than 20 years ago Davis (1964) investigated the impact of teacher preparation on the achievement test performance of secondary school students. The sample consisted of students of 18 chemistry teachers and 10 physics teachers. Using analysis of covariance, Davis found that students had higher adjusted test scores in chemistry when teachers had a master's degree rather than a bachelor's degree. However, students had higher adjusted scores on a standardized physics examination when teachers had a bachelor's degree rather than a master's degree. A number of weaknesses in the design of this study call into question the validity of the results. The small number of teachers in each subject area is a particularly serious weakness. Another problem is that teacher's length of service was not controlled and the possibility of an interaction between experience and advanced degree was not considered. In physics, for example, a young teacher could have more current knowledge of the field than an older teacher whose education is outdated even though the older teacher is more likely to have acquired a master's degree. Finally, since Davis could use only cooperating teachers, an unknown source of bias in sample selection may have influenced the results.

A more extensive investigation was conducted by Winkler (1975). The actual purpose of Winkler's research was to assess the effect of desegregation and student peer composition on student achievement. The results of his study, however, had implications for school and teacher effects research. Winkler first obtained estimates of student ability in terms of first grade reading achievement scores from student records. The primary outcome of interest in this longitudinal study was reading achievement score in the eighth grade. The school input variables included average teacher salary, student/teacher ratio and the proportion of teachers who obtained their undergraduate degrees from prestigious institutions. (Because the author noted that teacher salary was highly correlated with holding an advanced degree, we considered salary as a "proxy" for the degree-variable.) Separate regression analyses were conducted with 388 black and 385 white secondary school students in a single district in California. A significant positive relationship was found between student reading achievement in the eighth grade and teacher salary. Graduation from prestigious colleges was also positively related to student achievment. Similar results were obtained for both the white and black sample. One limitation of this study (for our purpose) was that the teacher salary variable was a function of both experience and possession of advanced degree, but unfortunately their separate contributions could not be estimated. Another limitation of this study was that data on the teacher variables were based on the "average of all teachers of verbal subjects in the grade, track and school of the student" (p. 194). Thus teacher characteristics were based on aggregated school level information and did not necessarily reflect on the actual teachers to whom particular students were exposed.

Despite the weaknesses noted, Winkler's study provided important contributions to the research literature on teacher/school effects because it used longitudinal student data; second, it involved replication across two different student samples; third, it was among the first to show a significant relationship between teacher characteristics and student achievement data after controlling for student ability. A minor criticism of this study might be that scores on the outcome measure were reported in terms of percentile rank. The use of standard scores might have been a better choice. Finally, the scope of the study was not as broad as if the author had examined the effects of school variables on achievement in multiple academic subjects rather than limiting his study to reading alone.

By contrast, Murnane (1975) attempted to develop a production function for student achievement in reading, mathematics and spelling. His samples were selected from predominantly black intercity elementary schools in a large metropolitan area. In his analysis Murnane considered seven teacher variables, four of which estimated the quantity and quality of teacher training. These included years of teaching experience, possession of a master's degree, undergraduate major (non-education or education), undergraduate grade point average, teacher gender, teacher race and teacher marital status. These data were collected on approximately 40 teachers from 15 schools. Children's standard test scores on the Metropolitan Achievement Test battery in reading, spelling and mathematics provided the output variables for the investigation. Murnane obtained data on two cohorts of children. The first cohort consisted of 440 third grade students and the second cohort included 442 students who were studied longitudinally in grades two and three. In addition to achievement data on the outcomes of interest, initial achievement data were available for the students from the previous school year. In all, 18 independent variables were entered into a single multiple regression analysis in Murnane's attempt to predict student achievement. Typically when so many independent variables are used in a single prediction equation, the results tend to be unstable from sample to sample. Murnane's study was no exception. His findings were inconclusive with respect to the effects of a teacher's master's degree on student achievement because of the inconsistency across samples for the magnitude, sign, and statistical significance of the regression coefficients associated with the master's degree variable. He also examined several interaction factors but none was statistically significant.

From our perspective a fairly serious problem with Murnane's study was that only a small number of teachers represented in this study seemed to have master's degrees. At best, in one sample, 25% of the children had teachers with a master's degree; in another sample as few as 7% of the children were taught by a teacher with a master's degree. From our examination of the data presented, 7% of the 442 students would be approximately 31 students (about the number in a single classroom), and thus it seems possible to assume that only one teacher with a master's degree was represented for this sample. Another criticism of this study of primary-grade children is that no consideration was given to the impact of kindergarten and first-grade teachers. These teachers' qualifications may have had some long-term influence on the performance of Murnane's second and third grade subjects, but there was no control for this.

Murnane's most stable finding was a positive relationship between teacher experience and achievement. His data suggested that students' achievement increased steadily as teacher experience increased from 1 to 3 years, but achievement declined slightly and stabilized as teachers' level of experience extended beyond 5 years. These results have sometimes been misinterpreted to imply that the effectiveness of a given set of teachers (or a single teacher) may decline as the teacher's level of experience increases beyond the three-year mark. Because Murnane did not study the same teachers over time, this conclusion is unwarranted. An alternate interpretation of results from this study would be that teachers' effectiveness tends to improve steadily over the first three years, but that after three years, there may be substantial attrition among the more effective teachers, while more of the less effective teachers remain in the classroom. (This possibility was noted by the researcher himself.) Considering that Murnane's focus was on black intercity schools, it would be not at all surprising if better, more experienced teachers sought teaching assignments in less-demanding school settings after three or four years. It is also possible that some highly effective teachers were promoted to administrative or special assignments, thus leaving the classroom, particularly if they had obtained master's degrees. Furthermore it seems likely that some of the more effective teachers may have left the teaching field altogether because of growing disenchantment with the job, the salary, or working conditions. In any case, this study raises several intriguing questions with regard to graduate training for teachers. For example, if beginning teachers had the benefit of master's level training at entry into their careers, could they produce levels of student achievement similar to those of the third year teachers who primarily held bachelors degrees? Or, if master's degree training were provided to teachers with 5 or more years of experience, could their effectiveness be increased? These possibilities are considered in conclusions of this chapter.

Another investigation into the effects of school variables on student achievement, using student as the unit analysis, was conducted by Summers and Wolfe (1977). The researchers randomly selected 103 elementary schools in Philadelphia then randomly selected 627 sixth grade students from within the schools. For each selected student, composite achievement grade equivalent scores were available from the third and sixth grades. In their analysis the change in achievement over the 3-year period was used as the school outcome of interest. Although the authors did not provide data in their report, the researchers claimed that the additional education beyond a B.A. degree for teachers was not related to student gains. Teacher characteristics for which the analysis was reported include a "quality" rating of the teacher's undergraduate college, teacher experience, and teacher score on the National Teacher Exam. The researchers analyzed the data with both the student as the unit of analysis as well as the data aggregated to the school level. The results differed slightly depending on the unit of analysis. With student data as the unit, a significant positive relationship was found between student gains and the "quality" rating of the teacher's college. However, a significant negative relationship was found between achievement gains and teacher scores on the National Teacher Exam. Later, however, when the data were analyzed using school average as the analytic unit, none of the teacher variables was significantly related to achievement gains. Given the design of the study, we believe the latter analysis was more appropriate. This study illustrates how tenuous are findings when student is selected as the unit of analysis.

Several methodological weaknesses in Summer and Wolfe's study include the use of gain scores (from third to sixth grade) as the outcome measure (rather than using third grade achievement score as an input variable in the regression); the use of grade-equivalents (rather than scaled or standard scores); and use of a composite achievement measure (for math and reading), rather than conducting separate analyses for mathematics and reading.

In summary then, we found a total of four large-scale studies, in which the relationship between teachers' possession of a master's degree and student's achievement was investigated using student as the unit of analysis. In one case, a positive relationship was reported for two separate samples between teacher salary and student achievement (and teacher salary was reported to be highly related to possession of the master's degree); in a second case a positive relationship was found for one sample and a negative relationship for another sample; and in the other two, no significant relationship was found. In each study, however, a variety of methodological problems threatened the credibility of the researchers' conclusions. Furthermore, as noted at the outset of this section we have reservations about use of individual student as the unit of analysis for assessment of teacher-effects. Thus, to date, no study has been conducted using student as the unit of analysis that allows us to draw definitive conclusions on this issue.

Studies Using Teacher (Classroom) As the Unit of Analysis

If the data are analyzed using classroom (or teacher) as the unit of analysis, then the question is: Does a class taught by a teacher with a master's degree typically have a higher mean achievement test score than a class taught by a teacher with a bachelor's degree? The policy implications of this question are not substantially different from when student is used as the unit of analyses, but results of the analysis are far more interpretable. When the goal of a study is to examine the impact of teachers' level of education on student performance, a strong argument can be made for using a research design that permits collecting the educational degree information from a sample of teachers and examining the educational achievement of their students in a way that permits direct linking between each teacher's educational status and the mean achievement performance of his/her class. Unfortunately, such studies have been reported only rarely in the research literature. Our search revealed four studies which addressed the question of comparative effectiveness of bachelor's and master's level teachers and used individual teacher (or classroom) as the unit of analysis.

One early study was conducted by Kleyle (1959). To examine the effect of variation in teachers' professional characteristics on teaching performance, she rated 108 elementary techers on the Beecher Teaching Evaluation Record. Kleyle found no significant differences in teaching performance which could be related directly to credits earned beyond the bachelor's degree or grades in student teaching.

Calabria (1960) conducted another study of the relationship between level of preparation and the teaching effectiveness but focused on secondary school teachers. While on the staff of the State Education Department, Division of Research in Higher Education, Calabria (1960) asked secondary school principals to nominate effective teachers of academic subjects. Over 1300 teachers were nominated and 770 agreed to participate. Five hundred twenty were sent postcard inquiries regarding their preparation and experience; 271 usable responses were obtained. Calabria reported that 86% of these effective teachers surveyed had a master's degree or its equivalent, and 67% had taken courses beyond the master's degree. In comparison, only 33% of all the teachers in the state of New York had received a master's degree at that time (Crane, 1958).

Calabria's findings in support of master's degree teachers are marred by several shortcomings in method. First, the criterion of teacher effectiveness was undefined. Principals were asked to nominate effective teachers but were allowed to use their own idiosyncratic criteria of effectiveness. Second, no comparison group was studied. Although there is a striking difference between the percent of teachers in the "effective" group holding master's degrees compared to the number having master's degrees in the total population of teachers, we cannot know for certain that a group of "ineffective" teachers would have necessarily differed from the "effective" teachers in the percent of teachers holding a master's degree. Finally we note that only 59% of the nominated teachers agreed to participate and only 21% actually completed the study questionnaire. This self-selectivity may have biased the final results.

Unfortunately neither Kleyle nor Calabria considered teacher effectiveness in terms of student achievement test performance. In 1973, however, Ober (1973) examined the relationship between teachers' characteristics and students' achievement gains on the reading and mathematics subtests of the Metropolitan Achievement Test. The sample consisted of 58 teachers and their 1,449 students from 11 elementary schools in a middle class suburb of a large midwestern city. Using multiple regression analysis, Ober found a significant positive effect on student achievement due to the interaction of the credits the teacher had earned beyond the bachelors degree and years of teaching. Specifically, as experience level of the teacher increased, the stronger was the positive effect of advanced training on their students' performance. (Note that this was one of the possibilities suggested by the results of Murnane's study discussed in the previous section.)

A recent promising study that could shed further light on this issue has been described by Peterson, Micceri, and Smith (1985). Their research effort centered around validation of the Florida Performance Measurement System using a sample of 468 elementary teachers, 226 middle school teachers and 528 high school teachers. In their publications, these authors note that data on teacher's degree were collected and were found to be unrelated to performance as measured by the FPMS; however, since this instrument was designed primarily for first-year teachers, it may not assess the types of behavior on which experienced teachers with or without advanced degrees might be expected to differ. Further work in progress described by D. Peterson (personal communication, 1985) involves collection of student achievement test data, but these results are not currently available.

In summary, studies of the comparative effectiveness of bachelor's and master's teachers, using classroom as the unit of analysis, have been relatively rare. In one early comparison using an observational rating scale Kleyle (1959) found no significant differences between bachelor's and master's level teachers. By contrast in a descriptive study Calabria (1960) found that among the "most effective" teachers (nominated by their principals) an overwhelming percentage of these teachers had completed master's work (while in the population of teachers in that state as a whole, only a small proportion had their masters degrees. In a more sophisticated study, Ober (1973) found that pupil mean achievement scores in math and reading were significantly increased with the combinations of teacher experience and educational credits earned beyond the bachelor's degree. No study was identified in which bachelor's degree teachers demonstrated superior performance to master's degree teachers. Thus when classroom has been the unit of analysis and teacher effectiveness is defined in terms of mean pupil achievement test scores or principal's nominations, the balance of empirical evidence tips modestly in favor of teachers with master's degrees.

Studies Using School/District As the Unit of Analysis

Studies in which school or school district served as the unit of analysis allow the researcher to address the question: Is the percentage of master's-level teachers employed in the district related to the average level of student achievement? When positive outcomes are obtained, results of such studies cannot be extrapolated to infer that a teacher with a master's degree will necessarily have a class with higher average achievement than a teacher with a bachelor's degree. Although this could be the case, it also could indicate that the master's degree teachers may exert a positive influence on curriculum, staff inservice programs, selection of qualified administrators, parental involvement, or other variables which contribute to overall student achievement throughout the school or district unit. In this case, the policy implication would be that hiring a high percentage of masters degree teachers is desirable for contributing to higher student performance, but that their influence may be beneficial to students beyond their own classrooms. Another important point is that in order for such studies to yield interpretable results, an effort must have been made to control for initial level of student achievement; otherwise positive results could simply mean that districts with more able student populations tend to hire and retain more highly educated teachers.

Most investigations using district as the unit of analysis have been of the type commonly characterized as "input-output" studies; school outputs typically include student achievement variables; school inputs commonly include teacher quality indices, class-size, school services, average district expenditure per pupil, etc.; student inputs often include student and family background variables (Glasman & Biniaminov, 1981). Nearly all of these studies have been conducted within the last two decades. In most cases, teacher's level of education was not the central focus of the study; thus this review includes both studies in which teacher's level of education was directly measured as well as some studies in which other variables (strongly related to teacher's education) were considered as "proxy variables" for teacher educational level.

When data are aggregated to the school or district level, there is considerable reduction in variation on the output measure. In other words, while there may be great differences in test scores at the individual student level within a single school or district, when the average test score for a school is used as the dependent variable, there may be relatively little variation among schools in a single district (or community). Similarly, because salary schedules and hiring policies are usually determined on a district-wide basis within a district, school-to-school variation in the percentage of teachers with master's degrees also may be relatively low. Such restrictions in variance on either the input or output variable decrease the chance of detecting a statistically significant relationship. On the other hand, studies in which data are aggregated to the district level (so that the average test score for a district is the dependent variable) and which include a broad sample of districts would be more likely to allow for sufficient variation to occur on the variables of interest. Thus in this section we first review studies in which school (or district within a single metropolitan area) served as the unit of analysis. Later we present a review of studies in which school or district was the unit of analysis and a broad geographic selection of districts was represented.

A common problem with many of the following studies arises from the use of stepwise regression analysis. In a widely used regression text, Pedhazur (1982) pointed out that in situations involving several intercorrelated predictors if one predictor has a slightly higher correlation with the criterion than the others have, in stepwise regression, not only will this predictor be selected first, but also there is a high probability that none of the remaining predictors will meet the criterion for entry into the model equation. It is erroneous, however, to conclude that the other predictors lack power to explain variance in the criterion. Studies of teacher characteristics using teacher salary, level of experience, and educational degree (which are highly correlated) in a stepwise regression are susceptible to this criticism.

Studies Within A Single District. One early study by Katzman (1971) focused on several school outcomes as a function of seven school characteristics and one community variable. Working with data obtained from 57 elementary school districts in Boston[1], Katzman developed separate regression models for predicting achievement in mathematics and reading. The reading outcome measure was recorded as the difference in median achievement between students in the second and sixth grades while mathematics was measured as the median achievment level of fifth grade students. Because the independent variables were correlated, the researcher attempted to identify the best subset of predictors by using stepwise regression. For prediction of reading performance, the subset of significant predictors included percent of experienced teachers, percent of students in classes with less than 35 classmates, and percent with fathers in white collar occupations. With mathematics as the outcome, the subset of significant predictors included the percentage of students' with fathers in white collar occupations, percent of teachers having a master's degree, percent of teacher turnover and age of the building. The last four variables were statistically significant and the last three were negatively related to achievement.

---

[1] Although this study involved multiple districts, their small size and location in a single metropolitan area accounts for its inclusion in this section.

The negative relationship between master's degree and mathematics achievement was an unexpected result which could not be explained by the researcher. There are, however, at least two reasonable explanations for this finding. One possible explanation arises from the analysis used by the researcher. In stepwise regression, the signs and magnitudes of the coefficients of variables which are entered into a model are not simply a function of the relationship between the predictor and outcome measure. Instead they are affected by the other predictors which have already been used in the model (Pedhazur, 1982). Another possible explanation arises from a different report of the same study (Katzman, 1968) in which the researcher considered outcome measures such as school average-daily-attendance, school membership (the percentage of students enrolled at the beginning of the year who remain throughout the year), school continuation (1.00 - dropout proportion); the percentage of students taking the statewide standardized Latin examination; and the percentage who pass the Latin examination. He concluded that for four of the six output variables, the percentage of teachers with master's degrees had a positive relationship. Perhaps the strongest effect of master's degree was in its relationship to school continuation, suggesting that teachers with master's degrees were more skilled at motivating lower-achieving students to remain in school than were bachelor's degree teachers. This is especially important since, by contrast, years of teacher experience was negatively related to school continuation rate. It may also help to explain why no relationship was found between reading achievement and teacher's degree. If the master's degree teachers were more successful at retaining potential dropouts in their classrooms, these students are likely to lower the overall mean test scores of those teachers' classes, thus making it appear that there were no differences in the achievement levels of students in bachelor's-level teacher and master's-level teacher classrooms. (The same explanation could account for the negative relationship between teachers' educational level and students' achievement in mathematics).

One problem with Katzman's (1971) study was that only one home background variable was included in the analyses. Katzman reported that other available data were considered but were not included in the final analysis because they were highly related to the percent of fathers in white collar occupations. A second problem with the study is that it was based on cross-sectional data obtained at a single point in time. Thus there was no control for student initial abilities and achievements. Also the gains in achievement were based on differences in median achievement between second and sixth grade students in the district. Since different students were involved in measuring gain, the results must be interpreted cautiously. Finally, the analysis was based on 57 districts, a relatively small sample size for the number of variables investigated. (Usually a ratio of 10 cases per variable is recommended for obtaining stable results). Taken together with the results of higher student retention rates for master's degree teachers (Katzman, 1968), the negative finding of the effect of teacher's master's degrees is highly suspect.

Burkhead, Fox, and Holland (1967) investigated the relationship between school inputs and student outcomes, replicating their study for three different community types but, unfortunately not always using the same variables. The researchers had obtained school level data from high schools in Chicago, Atlanta and a sample of high schools across the country from small communities (2,000-25,000 population) who were participating in Project TALENT. In Chicago the researchers examined mean 11th grade IQ and reading scores obtained from 39 high schools in the city. A school index was created by taking the ratio of the percent of students in the sample scoring in the 5-9 stanines to the percent of students in the normative group who scored in the 5-9 stanine. Nine school input variables and one family economic factor were considered. The teacher characteristics included as school inputs were median teacher experience and proportion of teachers with master's degrees or higher. The researchers determined order of entry of the predictor variables into a stepwise regression analysis. The first variable entered in all models was median family income.

In their initial stepwise regression analysis, the researchers made no attempt to control for student ability-level at entry into high school. From this analysis, they found that the only significant predictor of mean IQ score or reading score in the 11th grade was median family income; however, in a second analysis the researchers statistically adjusted for entry level ability by regressing their 11th grade IQ and reading scores on the 9th grade IQ scores and then analyzing the residuals. In this analysis, no teacher characteristic was found to be related to adjusted mean 11th-grade IQ score, but teacher experience level was a significant predictor of adjusted mean reading achievement score. A similar analysis was conducted using data from 22 high schools in Atlanta. In this analysis, however, the outcome variable was 10th-grade median verbal achievement score on the School and College Ability Test (SCAT). The predictor variables used in Atlanta differed slightly from those used in the Chicago analysis. Median teacher salary rather than experience or advanced degree was used on the basis of an arbitrary decision by the researchers after finding the three variables highly correlated with each other. After adjusting for 8th-grade median IQ scores of the schools, median faculty salary was not found to be a statistically significant predictor of verbal achievement for this sample. Finally, the researchers examined 12th grade mean reading scores of 177 small community high schools participating in Project TALENT. For this sample teacher experience was not a statistically significant predictor of 12th-grade mean reading score after taking into consideration mean school 8th-grade reading achievement. Thus in the first sample the significant contribution of teacher experience (which was correlated with teacher degree) may have prevented detection of a relationship between teacher degree-level and student achievement. In the two subsequent samples, teacher degree-level was not considered, but the "proxy" variables teacher salary and experience, were unrelated to student achievement.

A major contribution of the Burkhead et. al. investigation of school effects was their effort to examine school variables from several different geographic locations. The multiple site investigation provided some information as to the generalizability of the relationships between school inputs and outcomes. The results of the investigations indicated that the relationships between inputs and outputs within a single district may not be consistent across geographic regions. Among the major limitations of the study were the small number of schools included in the investigations in Chicago and Atlanta. A second serious limitation was the use of the stepwise regression procedure where the order of entering the variables was specified a priori by the researcher without a stated rationale. When the independent variables are interrelated, the order of entry is a major factor affecting the significance test of the variable. This poses a severe problem for those who are specifically interested in the effects of the percentage of teachers holding master's degrees. Namely, because master's degree and salary are highly correlated, once the variable of teacher salary has been entered into the prediction equation, the strength of the master's degree variable (as a predictor) has been substantially weakened. From the perspective of explaining variables in student achievement, since the master's degree and experience variables are responsible for teachers' salaries, rather than the other way around, it would be more sensible to enter these variables ahead of, or in lieu of, the salary variable. Furthermore, the preferred alternative analysis to stepwise regression would have been multiple regression with direct solution, since this type of analysis would allow the researcher to assess the contribution of each predictor separately to the output variable when the effects of all other predictors in the model are held constant. Finally, while the authors attempted to examine the "value added" by the school variables after controlling for previous achievement, the study was cross sectional which meant, for example, that data on reading achievement in the 8th grade was obtained on a different sample of students using a different test than the reading scores for 10th grade students.

Studies Across Districts. Thus far, the results discussed had indicated that school variables in general and teacher degree level in particular have little effect on student achievement. An argument can be made, however, that schools within a single district may be too homogeneous to permit identification of school effects. Many school variables (including teacher professional characteristics) are determined at the district level, and while similar within a district, may vary greatly across districts (Bidwell & Kasarda, 1975).

In the nineteen fifties and sixties three large-scale studies using school as the unit of analysis were reported. Mollenkopf and Melville (1956) studied the impact of 27 school and home background variables on seven different types of achievement scores. Mean achievement scores of 9th graders from a sample of 100 schools were analyzed in one phase of the study; mean achievement scores of 12th graders from 106 schools were analyzed in a second. Stepwise regression analysis was used. Percentage of teachers with 5 years or more of college training was considered but was not identified as one of the significant predictors of school achievement. A second large-scale study of about 3000 schools which received great national attention in the sixties was the study conducted by Coleman et al. (1966). The impact of 93 home, school, and teacher characteristics on 10 different achievement test scores was examined. After preliminary examination of correlations and regression equations, a set of seven teacher variables was selected by the researchers for further analysis. These included: teachers' SES, experience, degree level, teachers' score on a verbal ability test, and teachers' racial distribution. Considered as a block, this set of variables accounted for only a small proportion of variance (about 2%) in achievement test scores of white examinees and about approximately 8% of the variance in achievement scores for southern black examinees. Given the large number of variables included in the analyses and the high degree of relationship among the teacher variables, it is virtually impossible to assess the individual explanatory power of teachers' degree level in this study. Thus, neither the original Coleman study nor any of the subsequent reanalyses of these data by other researchers (Jencks, 1972; Mayeske, 1973, 1975; Smith, 1972) demonstrated a strong effect for teachers' degree level on student performance.

The most positive results in support of the importance of teacher degree level occur in a study by Perl (1973) who examined achievement of 3600 high school students sampled from the nationwide Project Talent sample. This large sample consisted of all students in a stratified random sample of 1000 high schools. Perl collected input data on each student's family background, peer-group background data, and a number of school variables. The impact of seven measures of teacher quality was considered, using percentage of teachers with master's degrees, percentage with Ph.D.'s, percentage of certified teachers, average years of experience, percentage of time spent in area of specialization, and average salary. The output variables were two different composite test scores. (The composites were derived from the two major largest principal components of a factor analysis of a battery of 22 separate tests.) The objective of Perl's analysis was to identify factors for which each $100 increase in school expenditure would correspond to an average increase of .8 - .9 percentile points on these two student output measures. One of the most effective teacher input factors identified was percentage of teachers with master's degrees. This is particularly noteworthy because years of experience and certification status were not significantly related to the achievement output of the schools. Teacher starting salary was also found to be significantly related to school achievement, as was percentage of teachers with Ph.D. degrees. From his production functions, Perl concluded that reallocation of school expenditure resources to increase starting salaries and to encourage teachers to attain the master's degree would have substantial payoffs in student achievement. One note of caution in interpreting Perl's results must be noted. In this study there was no direct control for student entry-level ability. Instead only student variables such as family income and father's occupation were directly controlled; while this is better than no control at all, we cannot be certain whether these variables serve as adequate "proxy" variables for student initial ability.

Bidwell and Kasarda (1975) directed another large-scale study of school effects using district level data as the unit of analysis. Data from 104 school districts in Colorado that enrolled over 90 percent of the students in the state were obtained. Focusing on the median percentile rank of secondary students in reading and mathematics, the investigators examined the relationship between these outcomes and five school characteristics including the ratio of pupils to teachers, the ratio of administrators to teachers, the percent of the certified staff with at least a master's degree, the ratio of professional support staff to classroom teachers and the percent of the population in the district which were non-white. The results of the analysis indicated that all of the input variables (except the ratio of professional staff to classroom teachers) were related to median achievement in reading. However, the percent of the staff having at least a master's degree was not significantly related to achievement in mathematics.

Bidwell and Kasarda's study was important for two reasons. First, they argued the issue of the appropriate unit of analysis for studying school effects. Second, the authors proposed a specific model of school effects and provided a theoretical rationale for the interrelationship between the variables involved. Unfortunately, however, they were unable to examine the "value added" by the school factor after controlling differences in ability across districts. Thus we cannot rule out the possibility that districts with more able students have more moster's degree teachers. Another limitation associated with district level data was that districts did not use the same standardized achievement test and the performance of different districts was based on different normative groups and on different test objectives.

A third investigation into school effects which used district level data as the unit of analysis was carried out by Brown and Saks (1975). These researchers used Michigan State Assessment data for fourth grade students in their analysis. Unlike Bidwell and Kasarda however, Brown and Saks argued that mean achievement data was an insufficient index to estimate school effects. They suggested that researchers should examine other distributional properties of achievement data. In particular Brown and Saks proposed the examination of test score variability. The researchers showed that school variables may not change the mean of the test score distribution but they could affect test variance. Separate analyses were conducted for city, (N=38), suburban (N=116) and town/rural (N=365) school districts across Michigan. The outcome under consideration was the average district composite achievement score; where the composite was the average of the reading, mathematics and mechanics of written English tests. Teacher characteristic variables that were used as predictors included average experience level, percent of teachers with a master's degree, and the ratio of students to teachers/administrators. For the mean achievement outcome, experience levels of teachers was a significant factor for both suburban and town/rural districts and percent of teachers with master's degree was also significant, but only for the town and rural districts. When the standard deviation of test scores in the districts was used as the outcome of interest, teacher experience was negatively related at a significant level for all three community types. Greater variability was associated with the percent of teachers having a master's degree in suburban districts but unrelated in both city and town/rural districts. This may indicate that in the suburban school districts, at least, master's degree teachers either taught students with more heterogeneous abilities or, more likely, that they were more successful in helping individual students achieve different levels of proficiency.

The major contribution to the school effects literature made by Brown and Saks' research was the inclusion of test score variability as an outcome to be considered when evaluating school inputs. The major weakness of the study was that it lacked control over student ability or previous achievement.

In summary, when schools within a single district, or district within the same city, served as the unit of analysis, the result has typically been that no significant relationship was observed between average level of school achievement and the percentage of teachers with master's degrees. Bidwell and Kasarda (1975) suggested that this may be due to the fact that schools within a district are fairly homogeneous in terms of their tendency to employ master's level teachers. When schools or districts represent a variety of geographic areas, studies conducted in the nineteen fifties and sixties using stepwise regression typically found no effect due to teacher degree level, but in the seventies several studies were reported in which percentage of masters' degree teachers was positively related to achievement mean or variance in the district. While this finding was not always true for every subsample or on every achievement subtest in each study, it occurred for at least one subsample in three of the large-scale multiple district studies conducted between 1973-1975 (Perl, 1973; Bidwell & Kasarda, 1975; Brown & Saks, 1975). In addition, it is important to note that percentage of teachers with master's degrees also has been found to be positively related to the proportion of students who continue school (as opposed to dropping out) (Katzman, 1968) and to greater variability in student achievement levels (Brown & Saks, 1975). Both of these outcomes seem to be at least as important as average school (or district) score on an achievement measure. Thus, while empirical findings at this level are mixed, we conclude that there is a general trend for districts which employ more master's-level teachers to have higher levels of student performance, as well as other positive educational benefits. This conclusion is somewhat tempered by the knowledge that in many of the studies reviewed, there was an imperfect attempt to control for initial level of student achievement. Although some attempt was made to control for initial level of student ability in nearly all studies reported here, the effectiveness of this control is somewhat uncertain in cross-sectional designs. This lessens our willingness to infer that the presence of greater numbers of the master's degree teachers in a district actually caused the higher levels of student performance.

Summary and Implications

In this review on the effectiveness of teachers with master's degrees we have considered three distinctly different types of studies. First were studies in which student was the unit of analysis; second were studies in which teacher and class mean served as the unit of analysis; finally were studies in which school or district mean was the unit of analysis. Results of these individual studies are summarized in Table 1.

Based on the fairly stringent statistical criterion used to declare that a finding is significant (α=.05), only 1 out of 20 studies is expected to show a positive relationship due to chance. In all we have reviewed a total of fifteen studies which provided data on the relationship between teachers' educational

Table 1

Summary of Effects of Studies Comparing Teachers with Bachelor's and Master's Degrees

Unit of Analysis	Study	Criterion Variable(s)	Findings
Student	Davis (1964)	Stand. Chem. Test; Physics Test	+;-
	Winkler (1975)	8th Gr. Reading Achievement Test	
		(a) black students	+
		(b) white students	+
	Murnane (1975)	Metropolitan Achievement Tests (reading, spelling, & math)	NS
	Summers & Wolfe (1977)	Composite Achievement Test Score	NS
Teacher/ Classroom	Kleyle (1959)	Beecher Teacher Eval. Rating	NS
	Calabria (1960)	Principal nominations of effective teachers	+
	Ober (1973)	Metropolitan Achievement Tests (reading, math)	+;NS
	Peterson, Micceri & Smith (1985)	Florida Performance Measurement System	Not yet completed

Table 1 (Cont'd.).

Unit of	Study	Criterion Variable(s)	Findings
School	Burkhead, Fox & Holland (1967)	1. IQ and standardized reading tests	NS
		2. Verbal ability on SCAT	NS
		3. Median reading scores	NS
	Katzman (1971)	Reading and mathematics achievement test scores	NS;-
		Composite achievement scores - verbal ability and abstract reasoning	NS;+
	Summers & Wolfe (1977)	Composite achievement grade equivalent gain	NS
District	Mollenkopf & Melville (1956)	Achievement test scores	NS
	Coleman, et. al (1966)	Achievement test score - verbal ability and abstract reasoning	NS
	Perl (1973)	Composite achievement score - verbal ability and abstract reasoning	NS
	Bidwell & Kasarda (1975)	Median percentile rank in reading and math	+;NS
	Brown & Saks (1975)	Composite score in reading, math, and mechanics of written English	+;NS;NS
		Variability of test scores	NS;+;NS
		(a) rural schools	+;NS
		(b) suburban schools	NS;+
		(c) urban schools	NS;NS

+ --Results favored teachers with master's degrees
NS--No significance difference was found
- --Results favored teachers with bachelor's degrees

degree-level and criteria of classroom performance or student achievement test scores. Some of these studies, however, used multiple samples and multiple outcome measures. In eight of these fifteen studies, for at least one of the samples studied, researchers found some evidence of a positive relationship between level of educational degree and one or more of the criteria. In two of these studies, a negative relationship was also observed between teachers' degree level and one of the criteria. In seven studies, no significant relationship was found between teachers' level of education and the criteria of student performance (as measured by test scores); however, in one of these studies teacher-education was found to be positively related to student performance and the method of analysis used may have obscured the effects of teacher performance. This nearly equal distribution of positive and non-significant findings appears to be similar across studies which used student, teacher, and school as the unit of analysis.

In terms of strength of research design, use of appropriate unit of analysis, and statistical methodology, however, we would rank the studies by Ober (1973), Perl (1973), Bidwell and Kasarda (1975), and Brown and Saks (1975) as the strongest studies. In all four of these, there was a positive relationship between level of teacher education and student achievement. Thus our final assessment of the limited empirical evidence is that it does provide a rationale for the current practice of encouraging teachers to seek professional training beyond the bachelor's degree and rewarding them for attainment of advanced degrees.

However, the data do not seem conclusive enough to warrant the suggestion that the master's degree should become the minimal level of educational attainment required to enter the teaching profession. All of the studies of this issue to date have involved an element of self-selection in terms of seeking a master's degree. Teachers who seek a master's degree voluntarily may differ in motivation, academic competence, or professional dedication from those who do not. Requiring a master's degree for all teachers might be less effective than improving the reward structure and professional recognition that accompany voluntary pursuit of graduate-level professional education. It may be that four-year baccalaureate preparation is quite adequate for students who are uncertain about career aspirations or who see teaching as a 3-5 year "transitional" occupation.

In addition, this review has suggested several other factors that are pertinent to the issue of employment of master's degree teachers. First is the problem suggested by Murnane's (1975) study; namely, that many effective teachers may be leaving the field after only a few years in the classroom. A salary schedule that rewards teachers who obtain the master's degree may in fact now offset this problem to some unknown degree. Another idea suggested from this review is that it may be those teachers who remain in the field (for perhaps three or more years) who could benefit greatly from graduate level work. This possibility is hypothesized from the findings of Murnane (1975) who found a decrease in student achievement to be associated with teacher length of service and Ober (1973) who found that increases in student achievement related to teachers' holding the master's degree was enhanced with teacher length of service. It may be that attainment of the master's degree helps the career teacher avoid or counteract the "burnout" syndrome. These ideas seem worthy of future study.

In addition, some emerging trends in teacher education raise new questions. Among these are:

1. How does performance of graduates of 5-year teacher-education programs compare with that of graduates of traditional 4-year programs?
2. How does performance of graduates of intensive, well-integrated master's programs compare with that of teachers who acquire their master's degrees through evening and summer coursework over an extended time period?

CHAPTER 3

DOES PROFESSIONAL EDUCATION MAKE A DIFFERENCE?

While one of the original questions guiding this review dealt with the comparative effectiveness of teachers who graduate from colleges of education and those who graduate from colleges of liberal arts and sciences, it became apparent almost immediately that few, if any, studies had addressed this question directly. A variety of studies, however, have been conducted which bear on this issue. These studies have been categorized as addressing one of the following broad questions:

1. Is there any evidence that graduates of colleges of education are more effective teachers than graduates of liberal arts colleges?

2. Is there a relationship between the number of college credits earned in professional education courses and teaching effectiveness?

3. Is there a relationship between number of college credits earned in a subject area and teacher effectiveness in teaching that subject?

The discussion of literature presented in this chapter is organized according to this framework.

Characteristics of Academic Institutions and Teacher Effectiveness

Three different types of studies have been conducted which should be considered separately incomparisons of education and non-education majors. These are studies of teacher's colleges, studies of different-size institutions, and direct comparisons of education and liberal arts majors.

Studies of Teachers College Graduates. Several early studies focused on comparisons of teachers who graduated from teachers colleges and those who graduated from 4-year colleges or universities. As noted in Chapter 1, there is a distinction between institutions traditionally devoted only to the training of teachers, called "teachers colleges," and colleges of education which are academic units within universities offering broader curricula. Schunert (1951) compared the effect of different types of teacher preparation on student achievement in geometry and algebra. The comparison focused on teachers who graduated from state universities, private colleges, and teachers college. One hundred schools were randomly selected from the population of 522 secondary schools listed in the Minnesota Educational Directory. The population was stratified by school size and administrative organization. From this sample, complete returns were obtained from 102 elementary algebra classes and 94 plane geometry classes, a return rate of 77%. (Using chi-square analysis, the author concluded that the teachers who completed the project were not significantly different in training and experience from the teachers who did not finish the project.) Students' mathematical achievement was measured by a locally developed test administered at the beginning and end of the year. The test was purported to measure (1) knowledge of mathematical concepts and principles, (2) mastery of mathematical skills, and (3) application of mathematical knowledge and skills to the solution of practical problems. The test was validated in a pilot study of six schools. Schunert found that the algebra achievement of students taught by graduates of state universities and private colleges was higher than that of students taught by graduates of teachers colleges. However, no difference was found in student achievement in geometry attributable to type of teacher preparation. In a similar study of 18 chemistry teachers and 10 physics teachers, Davis (1964) reported that students achieved more when their teachers had received the bachelor's degree from a liberal arts college than when their teachers had graduated from a teachers college. (For more detailed descriptions of this study, refer to Chapter 2.)

With the ultimate decline of teachers colleges, this issue now seems moot, but it is noteworthy that both of these studies provide empirical support for the present-day practice of educating teachers in the intellectual climate of a university setting.

Size and Prestige of Academic Institution. Some researchers have contended that characteristics such as size of the teacher's alma mater or its general academic prestige may be related to teacher performance. In this vein, Standlee and Popham (1958) investigated the effect of graduating institutions on teacher performance. The sample consisted of 880 teachers, all the 1954 bachelor's degree graduates from the 24 Indiana colleges and universities with teacher education accreditation. A single index of overall teacher effectiveness was derived from a "Teacher Ranking Form" that required principals to rank order their teachers with their peers. A higher proportion of graduates from intermediate size institutions (small public and large private in comparison to large public and small private institutions) were judged by principals to be higher in over-all teacher effectiveness. In addition, Standlee and Popham found significant relationships between the number of credit hours in professional courses and principals' ratings. However, Standlee and Popham discounted their significant findings, since only two of their twenty chi-square tests of the relationship between teacher preparation variables and teacher effectiveness were significant.

In addition several researchers have examined the effect of academic calibre of educational institutions on their graduates' effectiveness as teachers. The results show a general trend for teachers who graduate from prestigious universities to be more effective than those who graduate from institutions with less impressive academic rankings (Winkler, 1975; Summers & Wolfe, 1977). At least two interpretations of this finding are possible: (1) Graduation from a prestigious university may act as a proxy for teacher ability; or (2) teachers from prestigious institutions may be selected more often to teach in schools with high-achieving students. In any case, these studies raise the possibility that failure to control for a characteristic such as academic reputation of the alma mater may confound the comparisons of teachers who hold education or liberal arts degrees.

Comparisons of Education and Liberal Arts Majors. We found only three studies that were designed expressly to ascertain whether a bachelor's degree from liberal arts and sciences or a degree from a college of education is better preparation for a teaching career. Virtually all of the three comparison studies used principal ratings of teacher performance as the criterion. An important difference among the studies, however, was the degree of teacher experience. The first to be reviewed focused on experienced teachers; the second focused on first-year teachers; the third, on student-teachers. In the first study, Ellis (1961) examined relationships between teacher preparation of secondary social studies teachers and principals' ratings of classroom performance of these experienced teachers. In the fall of 1959 two groups of teachers were selected for comparison. One group, referred to as Group A, was made up of 44 teachers designated by their principals as "outstanding." The second group, Group B, consisted of 26 teachers considered as "average or below average" by their principals. The two groups did not differ significantly in the percent of teachers who had completed student teaching nor in the number of teachers who graduated from colleges of education. However, Group A teachers had completed a significantly greater number of semester hours in student teaching in social studies than Group B teachers. (Although it is not specifically stated, it seems reasonable that education majors probably tended to have more semester hours of student teaching than non-education majors.) In addition, Group A exceeded Group B in terms of college grade point averages in professional education, but the difference was not statistically significant. Since nineteen of twenty comparisons favored Group A, Ellis concluded that "patterns of variables involved in teacher preparation may be more clearly related to the professional performance of the teacher of social studies than is any individual variable."

Over a decade later, Copley (1974) studied three groups of beginning teachers in Missouri: (1) 22 liberal arts graduates with no professional education courses, (2) 38 liberal arts graduates with education courses but no student teaching, and (3) 40 bachelor of science in education graduates. The groups were stratified to include an equal number of different majors in each group. The principals rated the teachers from 0 to 3 on a 20-item rating scale, with 0 indicating that the teacher ranked below the 25th percentile of all teachers, 1 indicating ranking between the 26th to 50th percentile, 2 from 51st to the 75th percentile, and 3 indicating ranking above the 75th percentile. The outcome clearly favored graduates of colleges of education. Chi-square analysis favored the group of education graduates on the items: (1) exhibits understanding of people, (2) uses effective communication skills, (3) exhibits skill in managing classroom, (4) secures effective teaching results, (5) is considerate of pupils, (6) is fair in relations with pupils. The three groups did not differ on the measures of physical or emotional health or personality characteristics.

Recently Denton and Lacina (1984) again compared the supervisor ratings of the classroom performance and self-ratings of the morale of secondary student teachers majoring in education and student teachers not majoring in education. Fifty-five education majors were compared with 27 teacher certification candidates majoring in other colleges. The nonmajors completed

22 semester hours in education [general teaching methods (3 hrs.), educational psychology (3 hrs.), teaching field methods (4 hrs.), and student teaching (12 hrs.)]. Education majors completed five additional courses for a total of 34 semester hours including secondary education, early field experience, subject matter of teaching, preparation of instructional materials, and adolescent psychology. Three measures of teacher effectiveness were used: (1) The Evaluation Profile, a 28-item Likert scale measuring instructional competence (20 items) and personal and professional competencies (8 items). (2) The Curriculum Content Checklist, a rating of the student teachers' effectiveness in planning two curricular units, completed by the university supervisor, and (3) the Weekly Reflections Sheet, a self-report of how the student teachers allocated their time and a rating of their morale for the week. There were no significant differences between the majors and nonmajors on the variables of planning or morale. On the Evaluation Profile non-majors were rated higher than majors on all ratings of the use of duplicating and audiovisual equipment, but education majors scored consistently higher than non-majors on introducing and concluding lessons. Because the latter variables seem more important for teachers than skilled use of audiovisual equipment, we interpret these results as offering weak support for the superior performance of education majors.

In summary then, if we extrapolate that education majors are likely to have more earned credits in practice-teaching and other types of classroom experience than non-education majors (as reported by Denton & Lacina, 1984), the few studies located and reviewed show that education majors have been rated higher than non-education majors on diverse criteria such as "overall outstanding performance" (Ellis, 1961); communication skills, interpersonal skills, classroom management, and effective teaching (Copley, 1974); and introducing and concluding lessons (Denton & Lacina, 1984). The only criterion in which non-education majors were rated higher was in use of duplicating and audio-visual equipment (Denton & Lacina, 1984). Although the number of studies is small and the criterion of supervisor ratings is highly subjective, the existing body of research evidence provides some basis for encouraging aspiring teachers to choose education as their major discipline. An important consideration noted by Evertson, Hawley, and Zlotnik (1984) is that research showing even slightly greater effectiveness of graduates of professional education programs is evidence of the efficacy of professional education programs in compensating for the generally lower aptitude of college of education students that is so often reported (e.g., Weaver, 1979).

One reason that such marginal differences have been observed between education and arts and sciences majors may be that the differences in their preparation are more apparent than real. Given most university curricular structures, the first two years of university education are likely to be similar for students with either major. Furthermore in institutions which offer accredited teacher education programs, students of both programs must take coursework that meets state certification requirements.

Thus two students, one with an education major, the other with an arts and sciences major, may actually attend the same courses in methods of teaching and complete the student-teaching practicum under the same conditions. Both types of students would have similar coursework and contact with faculty in the college of education. The only differences may be in the type of electives they pursue or in the breadth of coursework taken. For example, the liberal arts major may be required to take two years of a foreign language, while the education major pursues electives in areas such as education, child psychology, or tests and measurement which are more relevant to the professional degree. Thus the actual degree the teacher receives may matter far less than the amount of coursework that is taken in professional education courses (typically offered by colleges of education). We turn next to studies addressing this issue.

Amount of Coursework in Professional Education

In an early descriptive study, Pisaro (1958) surveyed 190 school superintendents in Indiana to determine the reasons for dismissal of teachers. He received responses from 71 or 37% of the superintendents. The superintendents supplied examples of 196 unsatisfactory teachers and 168 superior teachers and described the behaviors, attitudes, and characteristics that contributed most to their judgments of effective and ineffective teaching. A total of 509 characteristics were supplied to describe the ineffective teachers compared to 831 descriptors of the effective teachers. Among the variables that discriminated between superior and unsatisfactory teachers were the amount of college training and amount of professional education. Specifically, teachers with less college training and less professional education were more likely to be dismissed from teaching than teachers considered by their superintendent to be examplary.

Somewhat more rigorous studies are those in which researchers examine the relationship between number of credits earned in professional education coursework and teacher or student performance. Let us first consider studies conducted at the elementary grade-levels. Hurst (1967) measured teacher effectiveness in terms of students' achievement on the Metropolitan Achievement Test. A random sample of third grade teachers in the Oklahoma City public schools during 1965-66 was selected for participation. However, data from only 55 teachers were usable, due to teachers' failure to return the questionnaire or missing MAT data. Hurst conducted a median test to determine if there were differences in teaching experience between the study group and the random sample of 100 teachers. He also calculated a t-test of average student gain to see whether there were achievement differences in the two groups of teachers. No significant differences between the two groups were detectable in experience or student achievement, so Hurst concluded that the teachers studied were representative of the population of teachers. Analysis of variance indicated no significant relationships between the number of teachers' credit hours earned in mathematics education and students' math achievement test scores.

Hice (1970) explored the relationship between characteristics of first-grade teachers and the reading achievement of their students. Forty first-grade teachers participated in the study. Procedures for sample selection were not described. Seven teacher characteristics were investigated: (1) number of years of teaching experience, (2) number of years of teaching experience in first grade, (3) number of reading courses the teacher had taken, (4) achievement motivation, (5) affiliation motivation, (6) progressivism, and (7) traditionalism. Students' scores on the Metropolitan Readiness and Achievement Tests for first graders were the criterion measures. Teachers were classified into four success categories on the basis of the mean adjusted end-of-year reading achievement scores. Hice found that the number of reading methods courses taken by teachers was positively associated with the achievement of female students. Use of categorical rather than continuous data may have reduced the likelihood of finding additional significant effects.

A second group of studies has focused on secondary level science achievement. Taylor (1957) investigated the relationships between science teachers' preparation, attitudes, and experience and their students' growth in achievement and interest in science. Teacher attitudes were measured by the Minnesota Teacher Attitude Inventory; pupil achievement was measured by the Essential High School Content Battery from the

World Book Company, and pupil interest was measured by the Occupational Interest Inventory from the California Test Bureau. Eighty-three teachers from grades 9 through 12 participated in the study; 42 of the teachers were fulltime and 41 were part-time. Teachers' preparation was measured by (1) the total number of semester hours of professional education and (2) the total number of semester hours in college science courses. The correlation between teacher semester hours in professional education and student achievement was not statistically significant, but Taylor also compared the science achievement of students whose teachers fell above the median on three of the four factors with the achievement of students whose teachers fell below the median on three of the four factors. He found that the science achievement of students whose teachers scored above these medians was significantly higher than the achievement of students whose teachers fell below the medians. Taylor concluded that professional education, science training, teacher attitude, and experience may contribute jointly to successful teaching and recommended that future studies should examine the interaction of two or more factors. One notable weakness of this study may have been in the content validity of the test. Classes included in the study included 25 general science classes, 26 classes of biology, 22 classes of chemistry, and 10 classes of physics in grades 9 through 12. It is unlikely that the test used to evaluate teachers' effectiveness was equally valid across these varied courses.

Perkes (1967-68) investigated the relationship between junior high science teachers' preparation, teaching behavior, and student achievement. The sample included thirty-two junior high science teachers, the entire population of junior high science teachers from the six junior high schools in a suburban California community. Half of their students completed the Sequential Test of Educational Progress: Science Test Level Three (STEP) and the remaining half of the students completed the Junior High School Science Achievement Test (JHSSA), a 100-item test of student recall of factual material that accompanied the science textbook used in the district. The number of teachers' credits earned in science education (methods) was significantly related to STEP test scores, a measure according to Perkes of application and interpretation. In contrast, this teacher characteristic tended to be negatively related to students' scores on the recall test. There was some evidence that the relationship may have been stronger for students with middle to high IQ than for students with low IQ scores. Teachers with more credits in science education had more frequent teacher-student discussion, more frequent student participation in laboratory exercises, used more hypothetical questions, and stressed principles and applications more often. Teachers with fewer credits in science education were more likely to lecture, conduct demonstrations for the class, and ask factual questions.

In a study of a broader scope Lawrenz (1975) also examined the effect of professional preparation on students'performance in science. She obtained a stratified random sample of 236 secondary science teachers from 14 states -- 84 biology teachers, 111 chemistry teachers, and 41 physics teachers. The initial response rate was 60% ; a follow-up of nonrespondents showed no difference between the respondents and nonrespondents on selected variables. A randomly selected class for each teacher completed the Learning Environment Inventory, the Test on Achievement in Science (compiled from the National Assessment of Science items), the Science Process Inventory, and the Science Attitude Inventory. A stepwise regression showed no relationship between the number of credits teachers had accumulated in science methods courses and student achievement. Again, lack of fit between the achievement test and the student's actual science curriculum is a weakness of this study.

Finally in an experimental study, Nelson (1978) studied the effect of methods instruction on the effectiveness of preservice teachers' science lessons, as measured by their students' achievement. Preservice teachers from two science methods courses were randomly assigned to an experimental (N=17) and a control group (N=16). The two groups were subdivided into a high GPA and low GPA group based on their GPA in the required eight semester hours of college service. Each preservice teacher taught the same three lessons on formulating hypotheses from Science: A Process Approach II (AAAS, 1975) to a randomly assigned group of fifth and sixth-grade students. The experimental group received 45 minutes of instruction presenting strategies to be used in teaching the three lessons. A six-item instrument from the Science: A Process Approach Module 78 was used to evaluate the effectiveness of the instructor. Analysis of variance indicated that the students of the preservice teachers who received methods instruction had significantly higher scores than the students whose teachers received no instruction prior to teaching the lessons. This study is of interest because it demonstrates that when the methods course content closely matches the curriculum that teachers will follow, it may be relatively easy to enhance student learning of that curriculum.

In summary, among the seven studies reviewed in this section, five resulted in identification of a positive relationship between amount of professional coursework and teacher effectiveness. One showed that amount of professional coursework distinguished between superior teachers and dismissed teachers (Pisaro, 1958); significant positive relationships between amount of teachers' professional education and students' performance on standardized achievement tests occurred in three studies (Hice, 1970; Perkes, 1967 and Nelson, 1978); a positive effect for the combination of hours of professional education, hours of science, and teacher experience and student achievement was reported in one study (Taylor, 1957); and no relationship between amount of professional coursework and student achievement as measured by a standardized test was reported for only two studies (Hurst, 1967; and Lawrenz, 1975). While each of these individual studies was subject to methodological flaws, in most of these studies there was some evidence to indicate that when teachers receive instruction in how to teach a subject, it has a positive impact in their students' learning of that subject. One point of particular interest is Perkes' (1968) finding that teachers with more credits in science-education used discussion and laboratory participation more and stressed principles and application more in their instruction while teachers with fewer credits in science education relied more upon memorization of facts from the test. This may, in part, explain how and why methods courses enhance teaching effectiveness.

Coursework in Academic Subject Areas

An ongoing issue in teacher preparation continues to be the argument concerning the desirable balance between professional education coursework and coursework in the subject-matter area. In the preceding section, the results of the literature review seem to imply that greater amounts of professional education are beneficial to teacher effectiveness. Yet more time on professional education courses leaves less time for coursework in academic subject areas in a typical undergraduate program of 4-years duration. Thus in this section, we turn to the question: Does amount of coursework in academic subject areas contribute to teacher effectiveness?

A number of studies described in the previous section on the effect of professional preparation also examined the relationship between academic preparation and teacher effectiveness. These studies will not be described again in this section. Only their results will be reported. Readers interested in the context of the studies can refer to Appendix A or the descriptions in the preceding section.

In an effort to increase the variance typically associated with principal ratings, Standlee and Popham (1958) developed a measure of teacher effectiveness based on principals' rank ordering of teachers with their peers. They found a significant relationship between the number of credit hours teachers had earned in academic courses and principals' rankings of their effectiveness but discounted this finding because it was only one of only two significant chi-square tests among a total of 22 tests of the relationship between various measures of teacher preparation and teacher effectiveness. When so many independent chi-square tests are conducted at the alpha level .05 on the same sample, this number of significant results would be expected to occur by chance.

Three studies were identified which explored the relationship between teacher preparation in mathematics and students' mathematics achievement. Smail (1959) compared teachers having two years of college education with those having four years of college. The sample consisted of 97 teachers of grades 4, 5, and 6 in the Sioux Falls South Dakota public schools. Smail found no difference in students' mean-gain in arithmetic attributable to the number of courses in higher mathematics the teachers had completed. Hurst (1967) examined the relationship between teachers' preparation and students' achievement on the mathematics subtests of the Metropolitan Achievement Test (MAT). Analyses of variance indicated no significant relationships between the number of credit hours the teachers had earned in mathematics and students' computations on problem-solving subtests. Furthermore there was a significant negative relationship between the recency of mathematics courses and student gain on Problem Solving and Concepts. Teachers with the most recent mathematics courses had the lowest student achievement gains. (We attribute this latter finding to be a function of teacher experience; i.e., teachers who had most recently completed their academic courses were the least experienced.)

Rouse (1967) attempted to investigate the cumulative effect of teachers' preparation on students' achievement at the end of the period encompassing the first five years, the first seven and the first nine years of elementary school education. Students' mathematics achievement in arithmetic fundamentals, arithmetic reasoning, and fundamental and reasoning combined were examined. A low negative correlation was obtained between teachers' college mathematics preparation and students' arithmetic achievement for the period from kindergarten through the middle of grade 6 and from kindergarten through the middle of grade 8.

Three studies have focused on the effect of elementary or junior high teachers' preparation on students' achievement in science. From a study of science achievement of fifth graders, Caruthers (1967) concluded that pupils whose teachers were experienced and prepared (having an average of 18 hours in science) had the greatest gain in achievement. Pupils with teachers who were inexperienced, but prepared, had the second largest gain in achievement and pupils whose teachers were experienced and non-prepared in science had the third largest gain in achievement. Pupils whose teachers were inexperienced and non-prepared (having an average of 8 hours in science) had the smallest gain in achievement. In a more recent study of science achievement of fifth graders, Thoman (1978) found no significant correlation between number of credits in science and students' achievement on the STEP Science Test. However, he did find that the number of semesters the teacher had taken in high school science was significantly related to students' achievement. This study had several serious design flaws. The small size of the sample (29 teachers randomly selected from the population of 5th grades in southeastern Wisconsin) may have contributed to the failure to find a relationship. Also, although the author indicated that "there was wide variation in mean gain from class to class," (p. 40); i.e., differences in level of class ability were not controlled.

Finally Perkes (1967-68) found that the number of credits junior high school teachers had earned in science did not relate to their students' science achievement as measured by the Sequential Tests of Educational Progress (Science) nor to their teaching behavior.

At the secondary level, Taylor (1957) contrasted the upper and lower thirds of the distributions in science subject matter for 83 teachers from grades 9 through 12. The mean number of hours in science for the lower third was 19, and the mean for the upper third was 75. The overall mean in science was 45.5 semester hours. The correlation between the number of hours the teachers had taken in science courses and student achievement was

.18, significant at the .10 level. Also Davis (1964) studied 18 chemistry teachers and 10 physics teachers. Using analysis of covariance, Davis found that students achieved more when teachers had no formal courses in physics rather than 10 or more hours, had attended National Science Foundation summer institutes and had more than ninety semester hours of science preparation rather than 50 hours or less. The number of semester hours of preparation in chemistry and mathematics was not significantly related to student achievement. Physics students had higher adjusted scores on a standardized physics examination when teachers had 10 or fewer hours of mathematics rather than 30 or more hours and had 100 or more semester hours in science rather than 50 hours or less. The number of semester hours taken in physics and participation in National Science Foundation summer institutes were unrelated to students' physics achievement. As noted previously, the small sample of teachers and unit of analysis (student) are critical shortcomings of this study.

In secondary social studies Ellis (1961) found no statistically significant differences in amount of social studies preparation between teachers rated by their principals as "outstanding" and those rated as "average or below" though more of the Group A teachers had declared majors in social studies. Group A exceeded Group B in terms of college grade point averages in social studies, but none of the differences were statistically significant.

More researchers have addressed the relationship of teacher preparation and student achievement in high school biology than for any other subject matter. Howe (1964) studied the relationship of teacher preparation and teaching methods in tenth grade biology classes in Oregon. Fifty-one teachers were selected for participation by stratified random sampling of Oregon schools. Howe found that none of the classes taught by teachers with less than 40 quarter hours of preparation in all science areas and less than 30 quarter hours in biology (with one exception) ranked in the upper third in gains in any of the five learning outcomes measured: (1) knowledge and understanding of biological facts, concepts, and principles; (2) skill in applying the methods of science; (3) improvement in critical thinking skills; (4) development of an understanding of the nature of science; and (5) development of more favorable attitudes toward science and scientific careers.

In three other studies the relationship between teachers' preparation in biology and their students' performance on the Nelson Biology Test was examined. Using multiple regression analysis, Sharp (1966) found a small but nonsignificant positive relationship between the number of semester hours of preparation in biology, chemistry, and physics and students' performance on the Nelson Biology Test. The number of semester hours of preparation in mathematics had a slightly negative relationship to students' biology achievement. In Osborn's study (1970), one-third of the students were taught by teachers having 16 or fewer hours preparation in the biological science, one-third by teachers with 17 to 32 hours of preparation in biology, and one-third by teachers with 33 to 48 hours of preparation. He found a non-significant positive relationship between teachers' preparation in biology and chemistry and students' achievement in biology. Osborn concluded, however, that the most effective biology teachers, as determined by their students' achievement scores, had taken a minimum of 12 hours in the biological sciences and a substantial amount of coursework in chemistry. Weaknesses in design of this study include the use of different groups of students for the pre- and post-test, and the use of only 8 students selected by random sampling as a measure of each teacher's effectiveness. Another limitation was use of a categorical measure of teacher preparation rather than a continuous measurement of the number od credits taken.

Culpepper (1972) randomly selected 18 teachers with 3-9 years of teaching experience from the 30 southern counties of Arkansas. His sample was stratified on the basis of the number of college credit hours taken in biology. The teachers were subdivided into three groups. Group 1 consisted of 6 teachers who had 16 or fewer college credit hours in biology, Group 2 consisted of 6 teachers with 17 to 32 credit hours in biology, and Group 3 teachers had from 33 to 48 credit hours in biology. Twenty students, were randomly selected from the teachers' classes. Culpepper obtained a significant correlation of .60 ($p<.05$) between the number of credit hours earned by the teachers in biology and the raw score gains of their students, but t-tests of the differences in the mean gains between the groups did not indicate a significant difference among the groups.

Rothman, Welch, and Walberg (1969) examined the relationship between physics teachers' preparation in subject matter, and their students' achievement and attitudes in science. Thirty-five male physics teachers who had volunteered to teach the new high school physics courses developed by the Harvard Physics Project comprised the sample. Student measures included (1) their scores on the Test of Understanding Science (TOUS), (2) their scores on the Physics Achievement Test (PAT), (3) their scores on the Welch Science Process Inventory (SPI), (4) the Tinkering subscore of the Pupil Activity Inventory, (5) the subscore on the Universe-beautiful and Physics-interesting subscores of a semantic differential measure of students' attitudes toward physics. The multivariate test of the hypothesis that there is no overall relationship between teachers' training, teaching experience, and knowledge of physics and students' changes in physics achievement, interest in science, and attitude toward physics was not significant; consequently, no tests of bivariate relationships were conducted. Results of this study could be questioned on the basis of the small size and select nature of the sample. Rothman (1969) replicated the study using a random sample of the national pool of physics teachers. Fifty-one teachers were selected randomly from a list of 17,000 physics teachers in the United States compiled by the National Science Teachers Association. Student learning outcomes were measured by the same variables included in the Rothman, Welch, and Walberg (1969) study. Canonical correlation analysis indicated an association between the teacher background and student learning variables. Zero-order correlations indicated significant relationships between the number of semester hours the teacher had earned in physics and the students' physics achievement (PAT) and their interest in physical science. The number of semester hours earned in mathematics was related to the students' TOUS scores, their physics achievement (PAT) and their interest in physical science.

Finally, in mathematics Soeteber (1969) investigated the effect of the number of semester hours of credit in mathematics teachers had completed on their students' scores on an algebra I test. The sample consisted of 34 teachers from 15 Wisconsin school systems. Teachers were divided into two groups. Teachers having 37 semester credit hours or more and those having 36 hours or less. Results of a two-way analysis of variance indicated no significant effect for credit hours alone; however there was a significant interaction between teacher knowledge and number of credit hours. Namely, among the group of 22 teachers who returned the advanced algebra test, students of teachers with fewer credits outperformed students of teachers with more credits in mathematics. The major limitations of this study are that over one-third of the teachers failed to return the knowledge test and the remainder took the test under unsupervised conditions. The credibility of these results is thus open to question.

It is difficult to draw any conclusions about the role of academic preparation on student achievement from the studies that have been conducted. They are fraught with methodological weaknesses that limit the likelihood of finding significant relationships. The studies examining the relationship of credits earned in subject areas and the effectiveness of elementary and junior high teachers yield no consistent results; that is, Smail (1959), Perkes (1967-68) and Thoman (1978) reported no relationship; Caruthers (1967) claimed evidence of a positive relationship; and Hurst (1967) and Rouse (1967) found evidence of a small negative relationship. At the high school levels, results of these studies are again equivocal. Davis (1964) and Ellis (1961) reported no relationship, but Taylor (1957) reported a small positive relationship. In the field of biology alone, Howe (1964) and Culpepper (1972) found evidence of a relationship, but Sharp (1966) and Osborn (1970) did not. Rothman, Welch and Walberg (1969) found no relationship between physics teachers' subject matter preparation and student achievement but Rothman's (1969) well-designed study of a random sample of physics teachers showed a clear positive relationship between number of hours in physics and students' physics achievement. In summary, only five of sixteen studies conducted showed a positive relationship between the number of credits teachers earn in academic fields and their teaching effectiveness. The majority of studies failed to support the hypothesis that increasing teaching subject-area preparation requirements will improve their students' performance.

Summary and Implications

The issue of the most appropriate preparation for teachers and the relative emphasis to place on preparation in educational methods or subject matter has been explored from three research perspectives in this chapter. The findings of this review are:

> 1. Comparisons of education majors and non-education majors, while few, have generally indicated that education majors are more highly rated by their supervisors than non-education majors. (See Table 2.)
>
> 2. When researchers have related the number of credits earned in professional education to student performance or supervisors' ratings, a positive relationship was reported in five out of seven of the studies. (See Table 3.)
>
> 3. When researchers have related the number of college credits a teacher earned in a subject area with student performance in that area, a positive relationship was found in only five out of sixteen of these studies. (See Table 4.)

Although in most cases the positive relationships reported have been for relatively small effect sizes and most of the studies suffered from methodological flaws, considered in concert, these findings seem to indicate that prospective teachers benefit at least as much, if not more, from their coursework in teaching methods as from preparation in academic subject areas.

This conclusion is supported by results of a literature review reported by Veenman (1984) who identified the most critical problems perceived by beginning teachers to be: classroom discipline, motivating students, dealing with individual differences, student assessment, parent relationships, inadequate instructional materials, and handling individual student problems. Inadequate knowledge of subject area was not among the major problems identified by either beginning teachers or their supervisors. One of Veenman's conclusions was that teacher preparation programs which accentuate acquisition of

Table 2

Summary of Results of Studies Comparing Teachers with Education Degrees and Teachers with Degrees in Other Fields

Study	Teacher Level	Criterion Variable	Findings
Ellis (1961)	Experienced	Principals' nominations of outstanding and below average teachers	NS
Copley (1974)	Beginning	Teacher rating scale (by principals)	+
Denton & Lacina (1984)	Student Teachers	Supervisor ratings of instr. competence; planning; self-reported morale	+;NS; NS

+ --Results favor teachers with education degrees
NS--No significant difference

Table 3.

Summary of Results of Studies of Relationship Between Number of Credits in Professional Education Coursework and Teacher Effectiveness

Study	Criterion Variable	Findings
Pisaro (1958)	Superintendants' nominations of unsatisfactory and superior teachers	+
Hurst (1967)	Metropolitan Achievement Test in 3rd-grade math	NS
Hice (1970)	Metropolitan Achievement Test in 1st-grade reading	+
Taylor (1957)	Essential High School Content Battery	NS(+)*
Perkes (1967-68)	Sequential Tests of Educational Prog. (Sci.) recall test of facts from test	+;-
Laurenz (1975)	National Assessment science items, Science Progress Inventory; Science Achievement Inventory	NS;NS;NS
Nelson (1978)	6-item test from a science curriculum module	+

+ --Positive relationship
NS--No significant relationship
* --A cumulative effect of professional credits in concert with other variables were observed.

Table 4

Summary of Studies of the Relationship of Amount of Coursework in Academic Subjects and Teacher Effectiveness in those Subjects

Study	Criterion	Findings
Standlee and Popham (1961)	Principals' ratings	NS
Smail (1959)	Gr. 4-6, Student gain in arithmatic	NS
Hurst (1967)	Metropolitan Achievement Test, math	NS
Rouse (1967)	Gr. K-6, Student achievement, math Gr. K-8, Student achievement, math	-;-
Caruthers (1967)	Gr. 5 Science achievement	+
Thoman (1978)	Gr. 5, Seq. Tests of Ed. Prog.	NS
Perkes (1967-68)	Jr. High, Seq. Tests of Ed. Prog.	NS
Taylor (1957)	Gr. 9-12, Science	+
Davis (1964)	Standardized chemistry and physics achievement test	-
Ellis (1961)	Principals' ratings of teachers	NS
Howe (1964)	Biology knowledge; scientific method application; critical thinking; understanding science; attitude	+

Table 4 (Cont'd.)

Study	Criterion	Findings
Sharp (1966)	Nelson Biology Test	NS
Osborn (1970)	Nelson Biology Test	NS
Culpepper (1972)	Nelson Biology Test	+
Rothman, Welch, & Walberg (1969)	Test of Understanding Science Physics Achievement Test; Test of Understanding Science; Welch Sci. Inv.; semantic diff. interest scale	NS
Rothman (1969)	same as above	+;+
Soetbeber (1969)	algebra I test	NS

+ --Positive relationship
NS--No significant relationship
- --Negative relationship

academic subject matter at the expense of the skills of instruction are justifiably subject to criticism. The general findings of our review support that conclusion.

CHAPTER 4

DOES TEACHERS' KNOWLEDGE OF SUBJECT AFFECT THEIR PERFORMANCE?

It seems reasonable to assume that how much teachers know about an academic subject should be related to their effectiveness in producing student achievement. In the pages that follow we examine the evidence on the relationship between teacher knowledge and student achievement for the purpose of determining the role of teacher knowledge in influencing student achievement. For ease of comparison, the studies have been grouped according to the method used to measure teachers' subject matter knowledge.

Studies Using the National Teacher Examinations

The most popular approach to studying the relationship between teacher knowledge and student achievement has been with the National Teacher Examinations (NTE). First administered in 1940, the NTE consists of the Common Examinations which yield a total score (WCET) that is a weighted combination of subtests in General Education and Professional education and examinations in the subject matter and methods of 24 areas. We describe first studies of the relationship between teachers' scores on the NTE and ratings of their classroom effectiveness.

The relationship between teacher knowledge and their principals' ratings of their performance was examined in several early studies. In a review of the validity of the NTE, Quirk, Witten, and Weenberg (1973) described a total of seven studies conducted between 1946-1969 in which correlations between NTE or WCET scores and supervisor ratings ranged from -.15 to .23 (i.e., Lins, 1946; Ryans, 1951; Delaney, 1954; Thacker, 1964; Walberg, 1967; and Carsen, 1969). Quirk et al. also cited studies by Flanagan (1941) and Shea (1955) which yielded correlations in the .40s and .50s. Kleyle (1959) rated 108 elementary teachers on the Beecher Teaching Evaluation Record and found a significant relationship between teachers' performance ratings and their scores on the NTE.

The relationship between teachers' scores on the NTE and student achievement has been examined in only a few studies. Lins (1946) reported a correlation of .45 between teachers' scores on the NTE and their classes' average residual gain on standardized achievement tests. However, only seven teachers drawn from five different subject areas were included in the sample. (For such a small sample the reported correlation is not statistically significant.) Lins obtained a correlation of -.302 between 26 teachers' NTE scores and rankings in which their pupils compared them relative to other teachers with whom they currently had classes.

Sharp (1966) investigated the relationship between teachers' scores on the NTE and student achievement in high school biology. There was a small positive but nonsignificant correlation between students' performance on the Nelson Biology Test and their teachers' scores on the Biology and General Science Teaching Area Examinations of the NTE and the scores on the Common Examinations of the NTE.

A more recent study of the relationship between secondary teachers' NTE scores in biology and their students' achievement also yielded nonsignificant results (Romano, 1968). Teacher knowledge was measured by the National Teacher Examination (Biology Area) and student achievement was based on their students' residual gain scores on the Cooperative Science Test-Biology (Forms A and B). A sample of 50 teachers was randomly selected from the group of 257 Biology I teachers who returned their NTE scores to the researcher following a letter of request sent to the 434 Biology I teachers teaching in South Carolina in 1977-78. The residual gain scores of the 35 classes with complete data ranged from 8.1 to -11.7. Forty-six percent of the class means indicated a negative residual gain. Romano reported a positive but nonsignificant correlation (r=.17) between teachers' NTE scores and their students' residual gain scores. A number of design problems existed in this study. There was inconsistency in the degree to which the teachers had taught the material covered in the criterion test. Nine teachers reported they had taught three of the five objectives, nine taught four of the objectives, and seventeen teachers had taught all five of the objectives. Also control for differences in student ability was lacking. Finally, sampling bias may have affected the results because only teachers who responded to the survey were included in the study.

The most recently published study of the relationship between teachers' scores on the NTE and student achievement in mathematics and vocabulary was conducted by Ducharme, Sheehan and

Marcus (1978). One hundred nineteen first grade teachers and their 1836 students comprised the sample. The Metropolitan Readiness Tests (MRT) Word Meaning and Number subtests were administered in September, 1973, and the Iowa Tests of Basic Skills (ITBS) vocabulary and mathematics subtests were administered in September, 1974. Stepwise regression analysis was used to examine the relationship between the teachers' scores on the WCET and the class-average raw scores on the ITBS subscales. The class-average MRT raw scores were entered first into the regression equations to control for initial differences in achievement. Teacher degree and years of experience were the next entries "to control for spurious teacher effects" (p. 135). WCET scores were significant predictors of both mathematics and vocabulary achievement. The scores accounted for 3% of the variance in mathematics achievement and 2% of the variance in vocabulary. In light of the small amount of variance accounted for by the WCET scores, Sheehan and Marcus concluded that "the NTE simply do not measure many of the aspects of teacher training that are important for effective classroom functioning as measured by pupil achievement tests." Furthermore, they found that the effect of the NTE scores on achievement was confounded with race and when the effect of race was controlled, the WCET scores were no longer significantly related to achievement in either mathematics or vocabulary.

In summary, studies in which NTE scores were used to define teacher knowledge have generally indicated no significant relationship between these scores and student achievement test scores. For the most part only very modest correlations (many non-significant) have been obtained between NTE scores and supervisor ratings of teacher performance.

Studies Using Other Tests of Teacher Knowledge

The relationship between teacher knowledge and student achievement has been investigated in a number of different subjects using specific subject-area tests other than the NTE to measure teacher knowledge. If we assume that the nature of the relationship between teacher knowledge and student achievement may differ depending on the subject matter being taught, the small number of such studies in each teaching field make it difficult to draw conclusions from these studies.

Reading. Several researchers have examined the relationship between teachers' knowledge about reading and their students' reading achievement. Clary (1972) examined the relationship of teacher personality, knowledge of reading, years of experience, and number of years since last reading course to pupil achievement in reading. The sample consisted of the 23 fourth-grade reading teachers and their students in a Spartanburg, South Carolina school district. The teachers completed the Edwards Personal Preference Schedule, the Inventory of Teacher Knowledge of Reading, and a questionnaire. Their students completed the Science Research Associates Achievement Series, Reading, as part of the regular school testing program during October and completed an alternate form of the test in March. Stepwise regression analysis indicated that the best predictors of students' reading achievement were teachers' personality (i.e., exhibitionism) and teachers' knowledge of reading.

A more complex relationship between teacher knowledge and student achievement was reported by Edelman (1973) who also used the Inventory of Teacher Knowledge of Reading to investigate the relationship between teachers' knowledge of reading and their students' reading achievement. The sample consisted of 200 teachers from grades 4 through 8 in Chicago, Illinois Public Schools. Pupil achievement was measured by pupils' standard scores on the Reading and Word Knowledge subtests of the Metropolitan Achievement Tests. Edelman found no relationship between teachers' knowledge of reading and pupils' reading achievement analyzed as two continuous variables. When analyzed categorically, Edelman found an interaction between the teachers' knowledge, the students' initial achievement status, and the skill area measured. The greatest percentage of students achieved high gains in reading when taught by teachers whose high, middle, or low reading-knowledge category corresponded to the students' high, middle, or low initial reading achievement status.

Mathematics. Two researchers examined the relationship between teachers' knowledge of basic mathematical concepts and their students' achievement in mathematics. In a study of 97 teachers in grades 4-6 in the Sioux Falls, South Dakota public schools, Smail (1959) found no difference in students' mean-gain in arithmetic attributable to teachers' understanding of basic mathematical concepts. However, Bassham (1961), like Edelman, found evidence of an interaction between teacher knowledge and student achievement in his investigation of the relationship between teachers' understanding of basic mathematical concepts and their pupils' mathematical ability. The sample consisted of 28 6th-grade teachers from an urban school district. Teachers completed the Test of Basic Mathematical Understanding (Glesson, 1948). Their 620 pupils took the California Achievement Test, Arithmetic, Form AA, in September, and Form BB in April. The California Achievement Test, Reading, and the Henmon-Nelson Test of Mental Ability were administered to the students in the fall. An Arithmetic Interest Inventory was administered in September and April. Students' pre-experimental period differences in arithmetic and reading achievement, mental ability, and interest in mathematics were controlled. Correlations between teacher scores of basic mathematical understanding and deviation scores of pupil gain indicated that the relationship differed depending on the students' intellectual ability. Teachers' scores on the test of basic mathematical understanding were not significantly related to duration scores of gain for the total group of students or for students whose ability scores on the Henmon-Nelson Test of Mental Ability were below the average score for the total group. However, there was a significant relationship between teachers' knowledge and the gain in understanding of students with above average intelligence.

An evaluation of an inservice education program, using a non-equivalent control group, suggested that differences in teacher knowledge of the course content in mathematics may not account for the differences in student achievement. Norris (1968) compared the mathematics achievement of 6th grade students taught by 18 teachers who attended for 2-1/2 hours per week for 14 weeks an inservice course covering the mathematical concepts taught in the textbook used in the district with the performance of students taught by teachers who did not attend the inservice course. The students of the treatment group had significantly higher scores than the control group students, when their scores were adjusted for initial ability, even though there were no significant differences in the teachers' scores on the criterion test. Norris speculated that the students' increased achievement was not due to the teachers' increased content knowledge but to some other factors possibly teachers' increased confidence or motivation.

Three studies have focused on the relationship betweeen teacher knowledge and students' achievement in algebra. Soeteber (1969) examined the relationship between teacher knowledge as measured by scores on an advanced algebra test and students' performance specially constructed Algebra I test. The research participants were 22 algebra I teachers who completed the Advanced Algebra Test and returned it by mail and their 1,184 students. There was a significant effect for teacher knowledge on student performance and a significant interaction between teacher knowledge and the number of semester hours credit the teacher had in mathematics. The students of teachers who scored high on the Advanced Algebra Test and had 37 hours or less in mathematics had higher scores on the Algebra I test. Soeteber also found that teachers with higher grade point averages had students with higher scores on an algebra achievement test than students of teachers with lower grade point averages.

In contrast, two additional studies failed to provide evidence of a relationship between teachers' knowledge of algebra and their students' achievement. Begle (1972) studied the relationship between teachers' understanding of algebra and their students' achievement for 308 teachers who had participated in the National Science Foundation Institute. Teacher knowledge was measured by two locally constructed algebra tests, one on the real number system and the second on groups, rings, and fields. Their ninth grade algebra students completed a mathematics inventory and a Reference Test for Cognitive Factors in the fall of 1970, and they completed an algebraic computation and a non-computation test in the spring of 1971. Stepwise regression analysis indicated that students' pretest scores predicted their scores on the two algebra posttests. Teachers' scores on the two tests of their knowledge of algebra were unrelated to their students' scores on algebraic computation. Teachers' understanding of modern algebra (groups, rings, and fields) was unrelated to their students' algebra scores. Teachers' understanding of the algebra of the real number system was unrelated to their students' algebraic computation skills but was significantly related to their understanding of algebraic concepts. However, the correlation was too low to be considered educationally important. Begle speculated that his failure to find significant relationships may have been due to the select group of volunteer teachers who participated in the study. He argued that there is probably a threshold effect such that there is a certain amount of knowledge teachers need to help students learn and beyond that minimum level, there is no relationship between amount of teacher knowledge and student achievement.

Concerned that Begle's sample was a highly select and bright group of teachers, Eisenberg (1977) attempted to replicate Begle's study with a more representative sample of teachers. Eisenberg sought participation of all the junior high Algebra I teachers in Columbus, Ohio. Ten teachers did not participate because their principals refused to allow their schools to participate, and nine additional teachers refused to participate. The remaining 28 teachers and their classes completed the same tests used in Begle's study. Like Begle, Eisenberg found no evidence of a relationship between teachers' knowledge and student performance. However, Eisenberg solved regression equations for 15 predictor variables. As the ratio of variables to subjects is quite high in this study, we must question the validity of the findings of this analysis. Consequently, given the methodological weaknesses of the Begle and Eisenberg studies, we are unable to determine whether the results are a function of the design flaws in the study or the subject matter itself.

Science. Three studies of the relationship between teacher knowledge and students' physics achievement arose out of the implementation of the Harvard Project Physics in the 1960s. In the first study, Walberg and Rothman (1969) secured the participation of 36 teachers from a group of 500 from across the country who volunteered to teach the new physics course, Harvard Project Physics. Teachers' knowledge was measured by 36 items from the unit tests of the Harvard Project Physics course, and student outcomes were measured by 17 different criteria: seven scales from the Classroom Climate Questionnaire, the Physics Achievement Test, the Welch Science Process Inventory, six subscores obtained from a semantic differential instrument constructed specifically to measure students' attitudes toward physics and three measures of students' participation in science activities. The measures were administered to students selected randomly so that gain scores were calculated on the basis of about one quarter of the students in each class. Post-test scores were adjusted for initial differences. An overall multivariate chi-square test of the multiple regression of the 17 criteria on seven independent variables (teacher achievement, prior student achievement, class size, class variation and the interactions of teacher achievement with each of the other variables) was highly significant. After the main effect of teacher achievement was partialled out, none of the other variables contributed significantly to the prediction of the learning criteria. Two significant zero-order correlations were obtained between teacher knowledge and the learning outcomes. Specifically, teachers with higher achievement had students with lower grades, and their students' ratings of the beauty of the universe was lower. Walberg and Rothman conjectured that "smart teachers may give lower grades because their own intellectual standards are higher. . .[and] because of their greater mastery of physics, may present the astronomy unit of the course rather abstrusely" (p. 256). In summary, they concluded that "higher teacher achievement is associated with rather trivial, negative effects on learning" (p. 256).

One source of weakness in Walberg and Rothman's findings was that teachers were not representative of the population of physics teachers. They scored higher on the physics achievement test than a national normative sample, which leads us to wonder if a ceiling effect may have restricted the range of relationships obtained.

Rothman, Welch, and Walberg (1969) examined further the relationship between physics teachers' knowledge, preparation in subject matter, and their students' achievement and attitudes in science. Thirty-five male physics teachers who had volunteered to teach the new high school physics courses developed by the Harvard Physics Project comprised the sample. Teachers' knowledge was measured by scores on the Test on Selected Topics in Physics, a 36-item measure of a wide range of topics in physics. The multivariate test of the hypothesis that there is no overall relationship between teachers' training, teaching experience, and knowledge of physics and students' changes in physics achievement, interest in science, and attitude toward physics was not significant. Consequently, no tests of bivariate relationships were conducted. The small size and select nature of the sample raise questions about the generalizability of the results of this study.

Rothman (1969) improved on the two preceding studies by randomly selecting a sample of 51 teachers from a list of 17,000 physics teachers in the United States compiled by the National Science Teachers Association. The design was identical to the preceding studies. Canonical correlation analysis indicated an association between the teacher background and student learning variables. Zero-order correlations indicated a significant relationship between teacher knowledge of physics and students' scores on the Test of Understanding Science.

Two researchers have reported a positive relationship between teacher knowledge and student achievement in science. Norris (1970) examined the relationship between teachers' knowledge of biology and their students' achievement. Thirty teachers who attended National Science Foundation Institutes at Ball State University during 1969 and 1970 participated. The teachers completed the Tennessee Self-Concept Scale, a knowledge test covering nine areas of biological knowledge, the Commission on Undergraduate Education in Biological Sciences Test, and their students completed the Differential Aptitude Test (DAT), a measure of ability, and the Processes of Science Test (PST), a measure of achievement. An index of teachers' proficiency was calculated by computing the mean of their students' scores on the Processes of Science Test and dividing the value by the mean of their students' Differential Aptitude Test. Norris found a significant positive relationship between the teachers' scores on the biology knowledge test and their teaching proficiency scores (which were a function of their students' performances on the DAT and PST).

Lawrenz (1975) also examined the effect of teacher knowledge on students' performance in science. Teachers' knowledge was measured by the National Teachers Exam in Science, the Science Process Inventory, and the Science Attitude Inventory. In a stepwise regression analysis teachers' scores on the Science Process Inventory emerged as a significant predictor of student achievement on the Test of Achievement in Science and on the students' scores on the Science Process Inventory. Further regression analyses completed for individual science courses suggested that the strength of the relationship between teacher knowledge and student achievement varied from class to class.

In contrast, Thoman (1978) examined the relationships between teacher knowledge of science and achievement of fifth graders in science. Teacher knowledge was measured by the STEP 1A Science Test. No evidence of a relationship was found. However, a number of weaknesses in design reduced the likelihood of discovering a relationship. There was no control for students' ability; the sample consisted of only 29 teachers, and multicollinearity of teacher knowledge with the significant predictors may have contributed to the failure to find a relationship. Although the author indicated that "there was wide variation in mean gain from class to class" (p.40), differences in level of class ability were not controlled. Also, data were not presented indicating the range of scores to determine if ceiling effects might have restricted the magnitude of the correlations obtained.

In summary, then when teacher knowledge is measured by administration of specific subject area tests, results have been somewhat mixed. However, the methodological quality of the studies varies considerably. We consider the strongest studies to be those conducted by Bassham (1961), Rothman (1969), Norris (1970) and Lawrenz (1975). Three of these four studies found a positive relationship between teacher knowledge and student achievement. The fourth found a significant interaction between teacher knowledge and student ability.

Using Grade-Point-Average as a Measure of Teacher Knowledge

Other researchers have measured teacher knowledge in terms of the teachers' grade point average. All but one of these studies used ratings as the measure of teacher effectiveness and, consequently, are weakened by the reliability and validity problems of rating scales. In spite of the restriction in range that is typical of rating scales, all the significant correlations reported in these studies, with only one exception, favored teachers with a higher GPA. Massey and Vineyard (1958) found positive correlations between teachers' grade point averages and each of the 15 criteria used to evaluate teaching success. The correlations, however, were low ranging from .10 to .38, with statistically significant relationships obtained only for subject matter mastery, competence in English expression, general culture and character standards, and ideals. The teachers' high average grade point average (2.9) and the negative skew of the ratings (mean general rating of 4.0) restricted the range of the variables, thus reducing the magnitude of the correlations. Massey and Vineyard noted the consistent pattern of relationship but concluded that they were not high enough for predictive purposes.

Hertz (1955) studied the relationship between the teaching success of first-year teachers and their undergraduate academic standing. One hundred fifty-seven teachers participated in the study. Their principals provided the measure of teaching success. Hertz found that teachers in the top 40% of their graduating class tended to have higher ratings, but he concluded that principal ratings were not sufficiently reliable to yield dependable data.

Maguire (1966) examined the relationship between principals' ratings of teachers' performance and the teachers' grade point averages in the following areas -- internship, overall, general education courses, professional education courses, major. For secondary teachers, grades in internship, teaching field, and overall GPA were related to principals' ratings of their effectiveness during their first year of teaching. Principals' ratings for first and fourth year elementary teachers were unrelated to any of the GPA variables. Principals' ratings of secondary teachers' fourth year of teaching were negatively related to their GPAs in general education.

Siegel (1969) studied the relationship between teachers' undergraduate GPA and their success as first-year teachers as measured by principal ratings on the Beecher Teaching Evaluation

Record. Pearson product-moment correlation coefficients were calculated for teachers' undergraduate GPA, professional education GPA, non-education GPA, and major field GPA. The results indicated that for the 393 elementary teachers, undergraduate GPA (r=.10) and education GPA (r=.11) were significantly correlated with principal-rated teaching success. For the 617 secondary teachers, principal-rated teaching success was significantly related to undergraduate GPA (r=.20), education GPA (r=.16), noneducation GPA (r=.13) and major field GPA (r=.18). (We note that these small correlations attained significance because of the relatively large sample sizes.)

Perkes (1967-68) investigated the relationship between teachers' grade point average and student achievement. Teachers' GPA in science was significantly related to students' STEP test scores (a measure of application and interpretation, according to Perkes) though negatively related to students' scores on the recall test. There was some evidence that the relationships may have been stronger for students with middle to high IQ scores than for students with low IQ scores. Teachers with higher GPA in science had more frequent teacher-student discussion, more frequent student participation in laboratory exercises, used more hypothetical questions, and stressed principles and applications more often. Teachers with lower GPAs in science were more likely to lecture, conduct demonstrations for the class, and ask factual questions. Unfortunately GPA was confounded with number of credits in science courses so the individual effects of these variables could not be assessed.

In summary, when teacher knowledge has been defined in terms of the teacher's academic grade point average, a small, but positive relationship between teacher knowledge and ratings of teacher effectiveness has been reported in most studies. In addition, Perkes (1967-68) provided evidence of a relationship between teachers' grade point average and student achievement that suggests teachers with higher grades may foster development of higher-level thinking especially for high-achieving students.

Summary and Implications

In a recent review prepared by the General Accounting Office (1984), the authors concluded that "research to date has failed to show a straightforward relationship between teachers' knowledge and the subsequent learning by their students in mathematics and science, at least for teachers in classrooms in the early 1970's" (p.34). Their conclusion was based primarily on the studies by Begle (1972), Eisenberg (1977), Lawrenz (1975), Wilson and Garibaldi (1976), and the three studies by Rothman and his colleagues (Walberg & Rothman, 1969; Rothman, Welch, and Walberg, 1969; Rothman, 1969). Our review does not support this conclusion. (See Table 6.) The Lawrenz study and the Begle study do show evidence of a relationship, albeit a small one. Furthermore the samples in the Eisenberg study and two of the Rothman studies were too small to yield dependable findings. In all, among fourteen studies of the relationship between teacher scores on a subject test and their students' achievement, only six yielded non-significant findings.

From our review we conclude that:

1. The National Teacher Examination score is not a good predictor of either teacher performance or student achievement (see Table 5);

2. Teachers' subject-area knowledge (as defined by GPA or subject area tests) makes a small contribution to classroom teaching behavior and student achievement (see Table 6);

3. The relationship holds most consistently for high-achieving students (Bassham, 1961; Edelman, 1973; Lawrenz, 1975; Perkes, 1967-68);

4. The relationship may be more relevant for students' achievement involving higher-order skills than for factual recall tests (Perkes, 1967-68). The meta-analysis of science teacher characteristics and student achievement by Druva and Anderson (1983) offers some insight as to why the relationship is stronger for higher-level learning. Two studies that examined the relationship between teacher knowledge and teachers' use of higher level, more complex questions yielded an average correlation of .36. The finding that teachers with greater knowledge ask higher level questions more frequently suggests that teachers with greater knowledge are more likely to foster students' understanding of complex scientific subject matter.

5. While teachers with greater academic knowledge may impart greater knowledge to their students, they are not necessarily more successful in creating positive attitudes toward the subject; nor are they always rated more highly by their supervisors;

6. Teachers' grade-point average tends to be a somewhat more stable predictor of teacher performance than teachers' scores on a single test.

In future efforts to develop a model of teacher effects on student achievement, the role of teacher knowledge should be considered. Several alternative relationships seem possible. On one hand teacher knowledge may directly influence student achievement. Another possibility is that teacher intellectual ability (specifically, verbal ability) jointly affects both teacher knowledge and teacher classroom effectiveness. A third possibility is that teacher knowledge may act as a moderator variable that interacts with the teacher preparation program or student characteristics to influence student achievement differentially in different circumstances. The research reviewed here on this topic to date does not preclude any of these possible alternatives.

Table 5

Summary of Studies of Teacher NTE Scores and Teacher Effectiveness

Study	Criterion	Findings
Flanagan (1941)	Supervisor ratings	+
Lins (1946)	Composite ratings	NS
Ryans (1951)	Supervisor ratings	NS
Delaney (1954)	Composite ratings	NS
Shea (1955)	Supervisor ratings	+
Kleyle (1959)	Supervisor ratings (Beecher Teacher Evaluation)	+
Thacker (1964)	Supervisor ratings	NS
Eissey (1967)	Supervisor ratings	NS
Walberg (1967)	Supervisor ratings	NS
Carsen (1969)	Supervisor ratings	NS
Lins (1946)	Standardized Achievement Test	NS
Sharp (1966)	Nelson Biology Test	NS
Romano (1968)	Coop. Science Test - Biology	NS
Ducharme, Sheehan, & Marcus (1978)	Metropolitan Reading Readiness; Iowa Tests of Basic Skills - Math	+ NS

+ -- Positive relationship
NS-- No significant relationship

Table 6

Summary of Studies of Relationship Between Teacher Knowledge and Teacher Effectiveness

Type of Criterion	Study	Findings
Performance Rating	Hertz (1959)	+
	Siegel (1961), Phase I	+
	Siegel (1961), Phase II	NS
	Maguire (1966)	+;NS
Student Achievement Scores	Massey & Vineyard (1958)	+
	Smail (1959)	NS
	Basshan (1961)	+
	Perkes (1967-68)	+;-
	Norris (1968)	+
	Rothman (1969)	+
	Rothman, Welsh & Walberg (1969)	NS
	Soeteber (1969)	+
	Walberg & Rothman (1969)	-
	Norris (1970)	+

Table 6 (Cont'd.)

Type of Criterion	Study	Findings
Student Achievement Scores	Begle (1972)	NS
	Clary (1972)	+
	Edelman (1973)	NS
	Laurenz (1975)	+
	Eisenberg (1977)	NS
	Romano (1978)	+
	Thoman (1978)	NS

+ -- Positive relationship
NS-- No significant relationship
- -- Negative relationship

CHAPTER 5

DOES TEACHER CERTIFICATION MAKE A DIFFERENCE?

The effect of teacher certification on teacher effectiveness is questioned periodically in educational research when the demand for qualified teachers exceeds the supply. To meet the increasing demand for teachers, school districts institute strategies that permit the hiring of teachers who do not meet all the requirements for regular certification. Such a period occurred in the fifties and early sixties. The current teacher shortage has once again raised the question of the relationship of certification to teacher effectiveness to critical importance. According to a survey of 1979-80 bachelor's degree graduates teaching in May of 1981, 56% of those teaching science and mathematics were not certified or eligible for certification in the field in which they were teaching. Further, 22% of all teachers and 26% in specialty areas were not certified (Plisko & Dearman, 1983 cited in GAO, 1984). A survey of 1,000 secondary school administrators in December 1981 indicated that administrators considered half of the newly employed science and mathematics teachers to be "unqualified" to teach science and mathematics (Shymansky & Aldridge, 1982 cited in GAO, 1984).

Traditionally, school administrators have preferred to hire teachers who meet all of the established certification requirements. The prevailing perception has been that teachers lacking these requirements are not adequately prepared to meet the responsibilities of teaching. For example, Shuster (1955) explored attitudes of principals, supervisors, and superintendents in Virginia toward teachers who held non-professional teaching certificates. The sample consisted of 179 teachers, 441 secondary school principals, 58 general and high school supervisors, and 88 superintendents. Approximately 75% of the supervisors and 66% of the principals believed that teachers holding the nonprofessional certificate required more supervision than teachers with the professional certificate. Ninty-one percent of the principals and 73% of the supervisors and all the superintendents preferred to work with professionally certified teachers.

To examine the validity of the widely held belief that certified teachers are more effective than teachers who fail to meet the requirements for state certification, educational researchers typically have compared the performance of regularly certified, provisionally certified, and uncertified teachers during the periods of teacher shortage. These comparison studies can be classified into the following three general categories, reflecting basic differences in research methodology:

1. Comparisons of teacher (or student) performance under naturalistic classroom conditions using measures that are not normally part of the certification or hiring process;
2. Comparisons of teacher (or student) performance under specific instructional conditions controlled by the researcher using measures specifically developed for that situation;

3. Comparisons of teacher (or student) performance using measures that are used statewide for certification of beginning teachers.

Studies of the first type were conducted predominantly in the 1950s and 1960s. A small number of studies of the second type occurred in the 1970s coinciding with the trend toward use of criterion-referenced testing in the classroom. In the 1980s studies of the third type have been conducted in concert with the introduction of statewide minimal competency assessments for beginning teachers.

Studies Under Naturalistic Classroom Conditions

In Oklahoma, in a study of the relationship between scholarship and first-year teaching performance, 62 teachers were rated by their immediate supervisors on a 5-point scale, with 5 representing the highest evidence of performance (Massey & Vineyard, 1958). The teachers who had completed the teacher preparation program leading to Oklahoma's standard certification received a higher mean (4.14) on the 5-point general performance rating than teachers who had only completed enough of the professional preparation program to receive a provisional certificate (3.65). Although Massey and Vineyard did not report a statistical test of this difference, it appears to be rather substantial considering that the ratings were negatively skewed and the mean rating for all teachers was 4.0.

A similar study was conducted three years later in New York by Lupone (1961), who compared the performance of provisionally certified and permanently certified elementary school teachers.

The sample consisted of 240 teachers in their first, second, and third years of teaching in selected school districts in New York State. A group of 40 provisionally certified teachers and 40 fully certified teachers with 1, 2, and 3 years of experience were compared. Principals in participating schools rated one provisionally certified teacher and one permanently certified teacher on seven areas of teaching behavior: (1) human relations, (2) preparation, (3) planning and management, (4) subject matter instruction, (5) parent-teacher relations, (6) pupil-teacher relations and (7) evaluation. Across the first, second and third years of experience the permanently certified teachers were more effective in five of the seven areas rated: (1) preparation, (2) planning and management,(3) subject matter, (4) pupil-teacher relations and (5) evaluation. In addition, during the second and third years, the permanently certified teachers were rated as superior to the provisionally certified teachers in instruction.

A number of studies grew out of the practice of issuing emergency certificates to meet the demand for teachers in the late 1950s in Florida. Beery (1962), Gray (1962), and Gerlock (1964) compared regularly and provisionally certified teachers in terms of their effectiveness as measured by ratings. Hall (1962) compared their effectiveness in terms of the achievement of their students.

Beery (1962) compared 76 first-year teachers who were issued provisional certificates because they lacked all or some of the prescribed courses with 76 fully certified, first-year teachers.

None of the provisionally certified teachers had completed student teaching; 34 had taken no professional education courses. and 42 had completed at least one course in education. To the extent possible, each provisionally certified teacher was matched with a fully certified teacher from the same school with the same teaching assignment. Sex, age, over-all grade-point average, college major, and school granting the bachelor's degree were also considered in matching teachers, but exact matching on all of the dimensions was not possible. Comparisons of group means on these dimensions indicated that matching was satisfactory on all but age and number of years since graduation. The provisionally certified group was somewhat older and farther removed from graduation. The teachers were observed five times, twice by professional educators, twice by other professionals, and once by a former school superintendent. The observers and teachers were not informed of the purpose of the study. The observers used a modified form of the Ryans (1960) Classroom Observation Record, a rating scale designed to measure teachers' friendliness, business-like demeanor, and enthusiasm. An additional set of ratings was developed to measure the teachers' use of appropriate teaching techniques and the teachers' overall effectiveness. Subgroups of teachers with similar assignments were compared on the five measures of teaching effectiveness. Subscore means were compared for nine groups of teachers. For each of the 45 subscore means, the difference favored the fully certified teachers, and for 25 of the comparisons the differences were statistically significant. When the subscore means were tested as a set for significant differences, only one of the sets failed to reach statistical significance. Recognizing that the differences may have been due to the differences in age and recency of graduation between the provisionally and fully certified teachers, Beery used three procedures - (1) comparing a subsample of teachers who were similar in age and recency of graduation, (2) correlational analysis, and (3) analysis of covariance. He concluded that the reduction in the differences between the provisional and fully certified teachers when corrections were made for age and years since graduation was not large enough to affect the statistical significance of the differences between the two groups.

Gray (1962) compared teachers holding Florida temporary certificates, Florida graduate certificates, and Florida post-graduate certificates in terms of the adequacy of their preparation as measured by (1) the teachers' self-evaluations, (2) their principal's evaluation, and (3) their scores on the Minnesota Teacher Attitude Inventory. The 2,407 first-year, white teachers in Florida during the 1954-55 and 1955-56 school years and their principals were asked to participate. The response rate was approximately 50%. The teachers in each group were roughly equivalent in the number of hours completed in general education; about two-thirds of those with temporary certificates had met the requirements in methods, foundations, and special education, but the majority lacked practice teaching. Gray found that the teachers' certification status was directly related to quality of preparation reported by principals and to

MTAI scores. The same trend also held for self-evaluations for teachers with graduate certificates compared to teachers holding temporary certificates but not for those holding post-graduate certificates.

Hall (1962) compared the effectiveness of fully certified and provisionally certified first-year teachers in language arts and arithmetic. The major difference between the two groups was that the preparation of fully certified teachers included student teaching. The sample included 38 elementary teachers from grades three, four, and five -- 21 provisionally certified and 17 fully certified teachers. Teacher effectiveness was measured by grade equivalent gain scores in six areas from the Stanford Achievement Test: paragraph meaning, word meaning, spelling, language, arithmetic reasoning, and arithmetic computation. Mental ability was also measured. Multiple regression analysis was used to estimate the effect of pupil IQ, teachers' grade point average in college, teacher's age, credits in professional education, and the teacher's score on the How I Teach test on students' achievement gains. Teachers' credits in professional education were associated with student gain in all six areas. Analysis of variance indicated that gains in spelling were significantly greater for the pupils of certified teachers than for uncertified teachers. Similar trends were noted for gains in scores on the paragraph meaning and word meaning subtests. Student IQ was significantly related to pupil gains in all six areas.

Gerlock (1964) examined differences between professionally and provisionally certificated secondary school teachers in administrators' ratings of (1) personal qualifications, (2) teaching skills, (3) relationships with others, (4) professional ethics, and (5) moral and social ethics and performance. The sample consisted of 341 secondary school teachers (grades 7-12) in either general science, social studies, mathematics, or English -- 201 professionally certified and 140 provisionally certified teachers who completed their first year of teaching during the 1960-61 academic year. Principals evaluated teachers on 5-point scales using the Teacher Evaluation form prepared by the Florida State Department of Education for the annual evaluation of all teachers as required by the Florida Statutes. Chi-square tests were conducted. Significant differences were found in favor of the professionally certified teachers on (1) general health, (2) teaching skills, (3) observing the confidentiality of students, parents and school personnel, and (4) professional ethics and performance.

A study by Shim (1965) has been cited as evidence that students taught by uncertified teachers scored higher on achievement tests than students taught by regularly certified teachers (Evertson, Hawley & Zlotnik, 1984). However, careful review of this study reveals that the classes of the uncertified teachers had a higher average IQ than the classes of the certified teachers, and when this was taken into consideration, the differences in the achievement of students of regular and professionally certified teachers were not significant. The design of this study is of interest because Shim attempted to investigate the cumulative effect of four teacher variables -

grade-point average, bachelor's degree, certification, and experience. Shim's rationale was that these teacher variables have not been shown to have a strong effect on student achievement when the impact of a single year with a teacher has been investigated. Shim believed that a stronger effect might be found if the students were exposed to a specific teacher characteristic over a number of years. To examine this possibility, Shim identified a homogeneous population of students from a semi-rural school district who had attended grades one through five in that district. Teachers were dichotomously classified according to four variables: having a GPA above or below 2.50, having a B.A. degree or not, being certified or not, and having more or less than 10 years of experience. Students were identified who had been taught for four years by teachers belonging to each of the dichotomous groups. When the difference in average IQ of classes was taken into account, the teacher characteristics did not influence student achievement significantly.

Two studies conducted in Georgia during the 1960s offer insight into the differences in motivation that distinguish provisionally certified from professionally certified teachers. Carter (1967) analyzed personality characteristics of beginning science and mathematics teachers. One hundred fifty-seven first year teachers of science and mathematics were selected randomly from the population of beginning teachers in Georgia during 1965-66 and 1966-67. Professionally certified teachers reported being more satisfied with their teaching skills than provisionally certified teachers, but provisionally certified teachers scored higher on a factor-analytically-derived subscale of the Pupil Observation Survey, indicating that students tended to find these teachers more interesting.

Further evidence of the differences in attitude that characterize provisionally and professionally certified teachers was provided by Bledsoe, Cox and Burnham (1967) who compared two groups of randomly selected provisionally and professionally certified teachers in science, social studies, English and mathematics at the secondary level and elementary teachers of grades 1-6 on a set of 33 self-report and classroom behavior variables. The professionally certified teachers obtained significantly higher ratings than the provisionally certified teachers on 11 of the 33 criteria of effectiveness included in the study. Specifically, professional teachers were rated by observers as more systematic and responsible, more skilled in the use of teaching media, more competent in nonspecific teaching behavior, and generally more competent than the provisionally certified teachers. In addition, the professionally certified teachers were more satisfied with teaching and with their preparation. At the end of the first year 36% of the professional teachers left teaching in comparison to 59% of the provisional teachers. At the end of of three years 56% of the professionally certified teachers remained in teaching in comparison to 31% of the provisionally certified teachers.

Perl (1973) included the variable of teacher certification as a measure of teacher quality in a large-scale input-output study of school effects. The sample was derived from the stratified random sample of 1,000 high schools that participated in the Project Talent survey of high school seniors in 1960. From the 26,000 male students who responded to the one and five year follow-up questionnaire administered by Project Talent, every fifth student was included in Perl's sample, but missing data reduced the sample to 3,600 pupils. Educational output was measured by two principal component scores obtained from a factor analysis of a large battery of aptitude and achievement tests. The first principal component appeared to measure general information and verbal ability, and the second principal component measured abstract reasoning. Certification did not relate to either measure of ability. However, the starting salary of teachers, and the time teachers taught in their area of specialization related to the measure of verbal ability and the percentage of teachers with M.A. or Ph.D. degrees and the percentage of time teachers spent in their field of specialization were related to pupils' scores on the abstract reasoning test. Since the measures of teacher quality are related, there is a possibility that the linear additive model used in the analysis substantially underestimated the impact of certification on student performance.

Studies Under Controlled Classroom Conditions

A study conducted by Popham (1971) to validate performance criterion-referenced tests dealt certification its most serious challenge. Popham prepared instructional objectives, teaching material and performance tests for three units in different subject areas. The subject areas were auto mechanics, electronics, and social studies. Popham's basic assumption was that the performance tests "at least ought to be able to discriminate between experienced teachers and nonteachers with respect to their ability to accomplish prespecified instructional objectives" (p. 109). After experiencing considerable difficulty in locating a school district willing to participate and further difficulties in finding inexperienced teachers, Popham located 28 paired instructors for an auto mechanics field test, 16 pairs for an electronics field test, and 13 pairs for the social studies field test. All of the experienced teachers held California teaching credentials and none of the non-teachers in the three comparison groups had any teaching experience or teacher education coursework. In the auto mechanic and electronics groups, the nonteacher was randomly assigned to teach one of the classes ordinarily taught by the experienced teacher. The participants received the instructional materials approximately two weeks prior to teaching. Students took a pretest, then received 9 hours of instruction from either the nonteacher or their regular classroom teacher, followed by the posttest. The social studies instruction lasted only 4 hours and teachers' classes were randomly divided into two groups, one of which was assigned to the nonteacher, who taught in a separate room in the presence of a credentialed substitute. Random assignment of students to instructors eliminated the need for the pretest.

Analysis of covariance revealed no significant differences in the test performance of students taught by the experienced and inexperienced teachers. Of the possible explanations of this failure to find differences attributable to teacher training, Popham concluded that "experienced teachers are not particularly skilled at bringing about prespecified behavior changed in learners" (p.115). However, he suggested that this finding reflects the failure of teacher preparation programs to train students to formulate instructional objectives and achieve them.

Popham's results, however, must be interpreted with awareness of several critical factors. First, the initial purpose of this study was to validate the criterion-referenced testing procedure that he had developed. The comparison between certified and uncertified teachers was made only because he initially assumed that students of certified teachers would learn more from their teachers' presentations. When this failed to occur, Popham chose to interpret the finding as meaning that the tests were valid but that his assumption about the superior teaching skills of certified teachers was unfounded. An equally legitimate interpretation of this finding would have been that the tests lacked validity or that the teaching materials were so complete that instructor qualifications did not matter. It is also critical to note that the uncertified personnel used in Popham's study presented their instruction in classrooms that were presided over by certified teachers and these certified teachers remained passively in the room during the instruction by the uncertified "guest" instructors. Thus Popham's findings cannot be generalized to situations where uncertified teachers must function independently in a classroom with complete responsibility for such things as classroom management, curriculum planning, dealing with discipline problems, motivating students, test construction, development of instructional objectives, or lesson planning. Since beginning teachers typically report that these latter aspects of their job are more difficult than presentation of subject matter (Veenman, 1984), Popham's study does not seem to address sufficiently the point of whether uncertified personnel can function as effectively as certified teachers in the classroom.

In direct response to Popham's (1971) study, McNeil (1974) compared the teaching effectiveness of certified elementary teachers with untrained elementary education students enrolled in a beginning course in teacher education. Nineteen experienced elementary teachers from kindergarten through grade six from three schools, a minority school, a middle socioeconomic school, and an upper-socioeconomic level school, and 19 education students participated in the study. Modeled on Popham's performance test, the strategy in this study was for the "teacher" to achieve a specific instructional objective - one of the following six tasks: (1) space relations, (2) rhythm, (3) number combinations, (4) phonetic rule, (5) folkways, or (6) divergent thinking. Two novice teachers and the experienced teacher were assigned to a common group. Children in the experienced teacher's classroom were randomly assigned to either one of the novices or their regular teacher. Lessons were 15 minutes long, taught simultaneously in the same room by the two novices and regular teacher. Chi-square analyses indicated that the pupils of the experienced teachers scored higher on the criterion tests and expressed more interest in the lessons than the pupils of the novices.

To investigate the possibility that the regular teachers' familiarity with their pupils may have contributed to the results favoring the regular teachers, McNeil compared the teaching performance with familiar and unfamiliar students of the 19 novices when they became student teachers. McNeil found that the acheivement of students was greater when the student teachers taught unfamiliar students, but the pupils were more interested when taught by a familiar teacher. McNeil concluded that his earlier study sugggested that with experience teachers are more able to produce both achievement and interest in their students.

Studies Using Certification Measures

In recent years, the negative conditions of teaching combined with the increase of opportunities for qualified women in traditionally male fields have reduced the number of candidates seeking careers in teaching (Schlecty & Vance, 1983). The emerging need for more qualified teachers especially in the fields of mathematics and science has led to renewed interest in the question of the relationship between certification and teacher effectiveness. Several recent studies reflect this renewed concern.

Cornett (1984) reported on four recent studies comparing fully certified and provisionally certified teachers. The first study compared scores on the Georgia Teacher Certification Tests for two groups: (1) teachers who had received a bachelor's or master's degree in a teacher education program and (2) teachers who had received a bachelor's or master's degree in an arts and science program but had not taken enough hours of professional education courses to be regularly certified. The sample consisted of teachers employed in the public schools in Georgia for the 1982-83 school year who had taken the Georgia Teacher Certification Tests in 1981-82 or 1982-83. All teachers had taught less than 4 years in Georgia. The Georgia Teacher Certification Tests were designed to measure teachers' knowledge of content in teaching fields as reflected in the curriculum in Georgia public schools. Based on combined data for 2 years, the mean score for all teachers was 79.2 and ranged from 78.4 for science teachers to 81 for social studies teachers. Provisionally certified arts and sciences graduates scored .7 of one point higher than teacher education graduates overall, but comparisons of the two were not consistent across all fields. Certified teacher education graduates scores .6 higher in mathematics and .6 higher in science. Differences between the groups were greatest in humanities and communicative arts. The arts and sciences graduates scored 2.6 points higher in humanities and 1.7 points higher in communicative arts. However, since the mean differences between groups were small (ranging from .6 for the overall score to 2.6 points for humanities) and no tests of the statistical significance of the differences were calculated, it seems likely that these differences are due to chance variations rather than meaningful differences. The percentages of each group who scored in five intervals of the score distribution also were quite similar. At the highest score range the distribution was 12% and 9% for arts and science and teacher education graduates, respectively, and 66% and 68% at the two lowest levels.

In a second study, Cornett (1984) compared scores on the National Teacher Examinations of Louisiana teachers holding regular certification with provisionally certified teachers. The sample consisted of all teachers receiving temporary certificates in Louisiana from July 1982 to July 1983 (N=89). Six held master's degrees. The number of credit hours in education earned by this group ranged from 0 to 36, with an average of 9.5 hours. The comparison group of 105 teachers was selected by random sample. Twelve had received master's degrees.

The Weighted Common Examinations (WCET) consists of a test in professional education and one in general education that includes written English expression, social studies, literature, and the fine arts, and science and mathematics. Teachers holding a temporary certificate scored higher (619) on the WCET than teachers with regular certificates (602). However of the 63 teachers taking the Elementary Education Area Test, the 21 teachers holding temporary certificates scored 23 points lower than the 42 teachers with regular certificates, even though those holding temporary certificates outscored those holding regular certificates by 40 points on the WCET.

In a third study, Cornett (1984) compared the classroom performance of Georgia teachers with regular certification and those holding temporary certification. The study included all teachers who were graduates of arts and sciences programs holding provisional certificates in the district during 1982-83 (N=21). Eighteen teachers taught at the secondary level; three taught elementary school. The teachers averaged 2 years of teaching experience. The comparison group of regularly certified teachers was matched with the temporarily certified on subject area and level taught. However, this group had an average of 5.2 years of teaching. The evaluation measure was a locally developed teacher evaluation system adapted from the statewide evaluation instrument for assessing beginning teachers. The instrument measured 10 competencies using 33 indicators. Scores on each indicator ranged from 1 to 5, with 4 or 5 indicating satisfactory performance. The 10 competencies were instructional planning, communication skills, instructional techniques, understanding of the subject, enthusiasm, and classroom management. The mean score for the provisionally certified teachers was 150 out of a possible score of 165, and the mean score of the matched sample was 158. Because the regularly certified group had 5.2 years of experience in comparison to the provisional group's two-year average, the difference in performance may be attributable to the difference in experience.

Finally, Cornett (1984) compared provisionally certified and regularly certified North Carolina teachers' scores on the National Teacher Examination and their classroom performance as measured by a statewide evaluation instrument, the North Carolina Teacher Performance Appraisal Instrument. The sample was composed of all teachers who held provisional certificates from 1979 to 1983 (N=191). A random sample of 348 regularly certified teachers was selected from the 21,000 teachers obtaining regular certification from 1978 to 1983. Of this sample, the districts responded with information on 292 teachers. Principals rated teachers on 33 basic teaching functions. Teachers were rated as below standards (2), meets standards (3), or above standard expectations (4). The mean scores for the evaluation showed no differences between the provisional and regularly certified groups. Because very few teachers received unsatisfactory ratings, the lack of variance in the ratings limited the possibility of finding significant differences between the two groups.

More recently Hawk, Coble, and Swanson (1985) compared certified and uncertified mathematics teachers. Thirty-six teachers, 18 out-of-field and 18 in-field, and their 826 students participated in the study. Teachers' effectiveness was measured in three ways: (1) student achievement, (2) teacher knowledge, and (3) professional teaching skills. Student achievement was measured by the Stanford Achievement Test (general math) and the Stanford Test of Academic Skills (algebra). Teacher knowledge was measured by the Descriptive Tests of Mathematics Skills, and professional skills were measured by the Carolina Teacher Performance Assessment System (CTPAS), a validated rating system of five teaching responsibilities: (1) management of instructional time, (2) management of student behavior, (3) instructional presentation, (4) instructional monitoring, and (5) instructional feedback. Students of certified teachers achieved significantly higher scores than students of the uncertified teachers on the general mathematics and algebra tests. Certified teachers scored significantly higher on the mathematics achievement and elementary algebra tests, but there was no significant difference between the two groups of teachers on the arithmetic test. The in-field teachers received significantly higher ratings on instructional presentation on the CTPAS.

Summary and Implications

Of all the studies conducted comparing certified teachers with teachers who had not met all of the requirements for state certification, all of the significant findings with the exception of Popham's study (1971) and Perl's (1973) studies favored certified teachers. Although significant differences were not found between the two groups on every variable on which they were compared, when differences were found, they favored the regularly certified teachers. Unfortunately most of the measures used in these studies have been fairly limited in their sensitivity to differences between certified and uncertified teachers. For example, whenever performance rating instruments are used, teachers' scores tend to fall at the high end of the scale and variability among teachers is small [see the discussion of the Massey and Vineyard (1958) study and Cornett (1984)]. Thus in summarizing results of studies reviewed in this chapter, we have separated findings of studies which examined measures of teacher knowledge, teacher classroom behavior, and student performance separately (See Table 7).

Furthermore, a number of design weaknesses may have hindered the detection of differences between certified and uncertified teachers. The criterion measure of effectiveness in five of the studies cited was a rating, usually by the principal. Rating scales have been vigorously criticized for their lack of reliability and validity (Medley & Mitzel, 1963; Rowley, 1975). Input-output studies of the sort conducted by Perl (1973) are unlikely to show significant effects for certification because a number of related variables are usually included in the analysis that reduce the likelihood of finding an effect except for the variables that are entered first in the equations. Studies such as those described by Cornett (1984) suffer from restriction of range since only teachers who had already scored above the minimum cutscore on the state certification examinations were included. If these minimum certification requirements had not been in force it seems likely that the observed differences between certified and provisionally certified teachers might have been even greater. Although the research evidence does not show large differences supporting the superiority of teachers with regular certification over teachers who have not met all the requirements, it is consistent in showing small differences favoring teachers' holding regular certification. Thus in spite of the weaknesses in design of various individual studies, the consistent results supporting the superiority of certified over uncertified teachers must raise doubts about the wisdom of hiring teachers who do not meet state certification standards. When Bledsoe, Cox, and Burnham's (1967) results showing the smaller attrition rate of regularly certified teachers and their greater job satisfaction in comparison to the provisionally certified teachers are added to the other evidence favoring regularly certified teachers, the advantages of hiring teachers who meet certification standards are clearly evident (Greenberg, 1985).

Table 7

Summary of Results of Studies Comparing Certified and Provisionally Certified Teachers

Type of Criterion	Study	Findings
Performance Ratings	Massey & Vineyard (1958)	+
	Lupone (1961)	+
	Beery (1962)	+
	Gray (1962)	+
	Gerlock (1964)	+
	Cornett (1984) (Georgia)	+
	Cornett (1984) (South Carolina)	NS
	Hawk, Coble, and Swanson (1985)	+
Teacher Attitude, Satisfaction, and Longevity in Field	Carter (1967)	+
	Bledsoe, Cox, & Burnham (1967)	+
Tests of Teacher Knowledge	Cornett (1984) (Georgia teacher test)	NS
	Cornett (1984) (National Teacher Exam) WCET; Elementary subest	-;+
	Hawk, Coble, & Swanson (1985)	+

Table 7 (Cont'd.)

Type of Criterion	Study	Findings
Student Achievement Scores	Hall (1962)	+
	Shim (1965)	NS
	Popham (1971)	-
	McNeil (1974)	+
	Perl (1973)	NS
	Hawk, Coble, & Swanson (1985)	+

+ --Positive relationship
NS--No significant relationship
- --Negative relationship

CHAPTER 6

METHODOLOGICAL ISSUES

The review of the relationship between teacher preparation and student achievement presented in this report is revealing. There have been surprisingly few studies on this issue, in spite of the popular assumption that teacher quality is vital to student achievement. Since the Coleman Report (Coleman et al., 1966) challenged this basic assumption, there has been increased interest in examining the relationship between teacher characteristics and student achievement, but the studies that have been conducted are so fraught with methodological problems that the results are of questionable validity (Shulman & Carey, 1984). In this section, we describe the major methodological problems that weaken these studies. Our intent is to identify these problems in order that they can be avoided in the research design proposed in this research. The methodological issues that must be addressed in the design of future studies are the following: (1) sample selection, (2) insufficient description of teacher characteristics, (3) the control of extraneous variables, (4) the limitations of cognitive measures, (5) ratings as a criterion of teacher effectiveness, (6) the stability of teacher effectiveness, (7) the appropriate unit of analysis, (8) the problem of multicollinearity in regression analysis, (9) linear analyses and interaction effects, (9) the stability of teacher effectiveness, and (10) the shotgun approach to data analysis.

Sample Selection

Bias in the selection of samples is a serious problem in the studies of teacher preparation and student achievement. The ethical requirement that teachers and principals must consent to their inclusion in a research study introduces the likelihood that some individuals will decline to participate. The pattern of refusal is not randomly determined. For example, in the Coleman Report (Coleman et al., 1966) one of the few studies that attempted to obtain a nationally representative sample, the researcher obtained only a 59% response rate. The pattern of nonresponse introduced a major bias into the analysis (Bowles & Levin, 1968). Large urban school districts were significantly underrepresented in the sample.

An additional problem in sample selection is the need to obtain a sample large enough to produce dependable results. The four studies conducted by Cornett (1984) for the Southeast Regional Educational Board represent the most recent example of policy studies based on inadequate samples. For example, to compare the effectiveness of teachers holding regular state certification with provisionally certified teachers, 21 uncertified teachers were compared with 27 certified teachers. Needless to say, the conclusions based on such small comparison groups cannot be assumed to be valid for larger, more representative groups of certified and uncertified teachers.

The problem of an adequate sample size can be easily resolved. Statistical techniques (Cohen, 1977) can be applied to determine the approximate number of individuals needed to yield dependable results. In contrast, the problem of sample bias is less easily solved. The need to obtain teacher consent leaves researchers vulnerable to nonrandom response patterns; consequently, the significance of the research must be carefully explained to potential participants, and special efforts must be made to insure the participation of reluctant individuals.

Inadequate Specification of Teacher Education Data

Research studies of teacher education have not defined teacher preparation variables in precise and consistent terms. Consequently, the findings are not comparable across studies, and policymakers are unable to draw inferences from these studies that can guide future teacher education policies. For example, in studies of teachers' level of education teachers are often categorized as having a master's degree or not having a master's degree. There is typically no effort to distinguish between types of master's degrees, for example master's degrees in education versus master's degrees in the subject area. Furthermore, the teacher who has completed almost all the coursework for the master's degree may be classified as holding only a bachelor's degree. Clearly, such teachers are more similar in educational level to master's level teachers than to bachelor's degree teachers. Analyses that treat educational level as a continuous variable by quantifying the number of credit hours teachers have completed in relevant coursework is likely to yield more interpretable results than have been obtained from previous studies that have treated educational level as a categorical variable.

To obtain precise, accurate, and specific educational data, researchers should obtain teachers' educational transcripts and include in their analyses the number of credit hours teachers have earned in each educational variable of interest.

Controlling Extraneous Variables

The input-output studies of school effects that have proliferated since the publication of the Coleman Report have identified a wide variety of variables that influence student achievement. To obtain a valid estimate of the effect of teacher preparation on student achievement, the effect of these extraneous variables must be controlled. The review of the input-output analyses of schools prepared by Glasman and Biniaminov (1981) identified the variables that have consistently shown a relationship to student achievement. Their review is helpful in identifying the variables that are most likely to influence student achievement and, consequently, must be controlled if we are to isolate the effect of teacher preparation on student achievement. In the sections that follow, Glasman and Biniaminov's review is used to identify the student and school inputs that should be controlled to eliminate effects extraneous to teacher preparation on student achievement.

Student's Family Background

All input-output studies of school effects of educational achievement have included student background as an input variable, because of its consistently strong relationship with achievement. Glasman and Biniaminov (1981) summarized the incidences of significant results for frequently used measures of family background: measures of family background have included family size, family income, family occupational status, family possessions, parental education, and family's educational environment, and in all cases the more favorable the family background, the higher student achievement. Family size was a significant predictor in 7 or 8 studies, "family income - 5 of 7, family occupational status - 7 of 13; family possessions - 5 of 5; parents' education - 9 of 13; and family's educational environment - 4 of 4" (Glasman & Biniaminov, p. 515).

Student Characteristics

Gender. The gender of students varies in its relationship with student achievement. There is a tendency for males to score higher than females on measures of verbal and nonverbal ability, mathematics, and general information, while females tend to outperform males on measures of composite achievement, spelling, student attitudes, high school completion, and continuation in higher education.

Kindergarten Attendance. Levin (1970) and Michelson (1970) reported a positive relationship between attendance in kindergarten and student achievement in reading and math and aspirations of achievement.

Student Over-Age for Grade. The three studies that have included student over-age for grade found it was significantly related to student achievment (Boardman et al., 1973; Levin, 1970; Michelson, 1970).

School-related Student Characteristics

Sociodemographic Characteristics. With one exception (Winkler, 1975) all of the studies that have examined the racial composition of the school (Bidwell & Kasarda, 1975; Boardman et al., 1973; Bowles, 1969; Hanushek, 1972; Perl (1973), Summers & Wolfe, 1977; Tuckman, 1971; Wiley, 1976; Winkler, 1975) have found the percent of white students positively associated with achievement.

Student Attendance Characteristics. Murnane (1975) found student turnover to be negatively related to reading achievement in black elementary schools. Coleman et al. (1966) found that it was negatively related to achievement in the North but positively related to achievement in Southern schools. Other measures of student attendance positively related to student achievement include days present (Murnane, 1975) and quantity of schooling (Wiley, 1976). Summers and Wolfe (1977) obtained a negative relationship between number of unexcused absences and lateness and a composite measure of student achievement.

Prior Level of Student Achievement. Haertel (1986) has identified initial level of student competence in a subject as one factor which must be controlled before attempting to evaluate teachers in terms of student ahcievement. For example, Katzman (1971), Ober (1973) and Summers and Wolfe (1977) attempted to control for students' prior level of achievment by using gain scores. Anther common approach has been to enter prior achievement or aptitude as one of the variables in a regression analysis (exemplified by Burkhead, Fox, & Holland, 1967).

Student Attitudes. Student attitudes have been included as both inputs and outputs in studies of school effects. Internal control has been positively related to student achievement in five studies (Bowles, 1970; Boardman et al., 1973; Coleman et al., 1966; Hanushek, 1972; Levin, 1970). Bowles (1970) found a positive relationship between students' self-concept and students' academic operations and achievement. Mayeske et al. (1973, 1975) concluded that student attitudes were stronger determinants of verbal achievement than socioeconomic inputs.

School Inputs

School Conditions. Glasman and Biniaminov (1981) included three sets of variables in the category of school conditions: services, expenditures, and staff. Among the variables included in services, tracking was the only consistent predictor. In both Bowles's (1969) and Michelson's (1970) studies, tracking was negatively related to achievement. The effect of the number of books per student on achievement was positive in three cases, mixed in one, and negative in another. Class size was negatively related to achievement in 6 cases and positively related in 5. Regarding school facilities, science labs were positively related to verbal achievement (Bowles, 1970; Bowles & Levin, 1968); age of buildings was negatively related to achievement in four cases, and had mixed results in one instance: size of school site was positively related to achievement of elementary studies (Guthrie, et al., 1971; Michelson, 1970), and size of school enrollment was negatively related to achievement in four instances and positively related to the number of dropouts, continuation in higher education, and educational aspirations. Of the variables included, the category of expenditures, library expenditures were negatively related to composite achievement in elementary schools (Kiesling, 1969) and the number of dropouts in secondary schools (Burkhead et al., 1967); materials and supplies expenditures were also negatively related to the number of secondary school dropouts (Burkhead et al., 1967); administrative expenditures were positively related to composite achievement in elementary and secondary schools (Kiesling, 1969; 1970); instructional expenditures were a positive predictor of composite achievement in secondary schools and reading in elementary schools (Benson, 1965; Goodman, 1959). Extracurricular expenditures were positively related to verbal ability in secondary students (Cohn & Millman, 1975); total expenditures were positively related to achievement in secondary schools (Bidwell & Kasarda, 1975; Perl, 1973). In sum, expenditures were positively related to school output in every instance except library expenditures in the Kiesling (1969) study. Of the staff variables, administrative manpower was negatively related to reading and mathematics achievement in secondary schools and positively related to verbal achievement (Bidwell & Kasarda, 1975; Cohn & Millman, 1975); auxiliary manpower was negatively related to verbal achievement and self-concept in secondary schools (Cohn & Millman, 1975); teacher turnover was a positive predictor of nonverbal ability and reading achievement, a mixed predictor of verbal ability, and a negative predictor of mathematics achievement and educational operations. Teachers' salary was also an inconsistent predictor.

It had a negative relationship to student attitudes and the dropout rate, a positive relationship on verbal, mathematics, and composite achievement, grade point average, and interest in school.

Teacher Characteristics. A number of teacher variables have exhibited a relationship with student achievement. Teaching experience was a significant predictor of student achievement in 12 of 16 studies. Teachers' verbal ability has been a significant predictor in 7 of 8 studies, undergraduate instruction was significant in 4 of 11 studies, race was significant in 2 of 5 studies, and sex was a predictor in 2 of 3 studies. Teachers' teaching load was negatively related to the verbal and reading achievement, interest in school and self-concept of 11th graders (Cohn & Millman, 1975), and teachers' job satisfaction was positively related to verbal, reading, and mathematics achievement, students' grade aspirations, and interest in school (Levin, 1970; Michelson, 1970; Guthrie, 1971; Cohn & Millman, 1975). Teachers' sense of efficacy, that is, the extent to which teachers believe that they have the ability to teach and their students have the ability to learn, was positively related to student achievement in all 4 of the studies that have examined the relationship (Armor et al., 1976; Ashton & Webb, 1986; Berman et al., 1977; Gibson & Dembo, 1984).

In this section, we identified a large set of variables that might directly affect teacher effectiveness or might mediate the relationship between teacher education variables and student achievement. In summary, the variables that should be controlled in input-output studies include (1) students' family background, (2) student characteristics, including gender, kindergarten attendance, and students' average-age for grade, (3) school-related characteristics, including the school's sociodemographic characteristics, students' attendance, prior level of student achievement, and student attitudes, (4) school inputs, including services, expenditures, and staff, and (5) teacher characteristics, including teaching experience, teachers' verbal ability, teacher race and sex, teaching load, and level of motivation or sense of efficacy.

Limitations of Cognitive Measures

The most common criterion in studies of teacher effects is student achievement on standardized achievement tests. In a comprehensive review of input-output analyses of schools, Glasman and Biniaminov (1981) reported that 60% of the studies used only cognitive measures of output. Although the studies varied in the standardized achievement tests used, all the standardized achievement tests were norm-referenced and measured basic curricula. The use of such measures as the sole criterion of effectiveness ignores the fact that educational outcomes include a variety of important noncognitive as well as cognitive outputs that may vary in their relationship to educational inputs. The use of multiple outcomes reveals that such differential effects may require decisions regarding the relative importance of those outcomes. For example, Katzman's (1968) results suggest that an increase in the percent of teachers holding master's degrees would result in better attendance and higher aspirations but declines in mathematics scores. To use Katzman's data to determine school district hiring policy, the school district would have to decide on the relative importance of attendance, aspirations, and mathematics achievement to the community. Schofield (1981) reported further evidence that teacher characteristics may be differentially related to cognitive and noncognitive outputs. Fifty-six beginning teachers in Australia in grades 4 to 6 who had taken tests measuring their mathematics achievement during their last year of training administered tests of mathematics achievement and attitudes toward mathematics to all their students at the end of the first term and again at the end of their second term of teaching. The students of high-achieving teachers had the highest performance on both the mathematics concepts test and mathematics computation test at the end of both terms; however, these students had significantly less favorable attitudes toward mathematics than pupils of low- and middle-achieving teachers.

An additional problem with limiting the measurement of teacher effectiveness to the use of standardized, norm-referenced achievement tests is that such tests are "biased against finding large differences between schools in achievement," and consequently, "continued use of these kinds of tests in education will continue to provide biased evidence against any educational treatment effect" (Carver, 1975, p. 78). Thus, the traditional standardized tests used to evaluate teacher effects may lack the sensitivity necessary to reveal relationships between teacher characteristics and student achievement. In addition, Glasman and Biaminov (1981) pointed out that because disadvantaged populations tend to be underrepresented in the norm groups for these tests, the achievement tests are less valid for such groups.

To overcome the limitations of standardized norm-referenced achievement tests, researchers should develop multivariate evaluation measures that are matched to the content of the curriculum and validated in a number of different contexts (Dunkin & Biddle, 1974). In addition, noncognitive measures of effectiveness should be included, for example, students' attitudes toward school and the subject matter, absentee rate, and disciplinary actions.

Ratings as a Criterion of Teacher Effectiveness

Ratings have been the major criterion of teacher effectiveness in educational research on teacher education. Unfortunately, ratings have serious weaknesses that threaten their reliability and validity. Ratings are especially prone to bias as a result of the *halo effect*, the tendency to rate an individual consistently on the basis of a general impression (Kerlinger, 1973). For example, a principal may rate a teacher higher than the teacher deserves because the principal likes the teacher or because the teacher has been particularly supportive of the principal's policies. Thus, the rating of one chracteristic may unduly influence the ratings of other characteristics.

Rating scales are particularly susceptible to personal bias errors (Vockell, 1983), the tendency to rate everyone either high, in the middle or low. Teacher effectiveness ratings tend to be especially susceptible to the error of central tendency. This tendency to avoid extreme judgments by rating down the middle of a rating scale (Kerlinger, 1973) reduces the variability in scores thus limiting the possibility of finding relationships between teacher preparation variables and teacher effectiveness.

To avoid the threats to reliability and validity that weaken rating scales, researchers should obtain more objective measures of teacher effectiveness, for example, systematic observation data and student achievement test scores.

The Stability of Teacher Effectiveness

The search for relationships between teacher preparation and student achievement is based on the assumption that teacher effectiveness is a relatively stable characteristic, and research on effective teaching has generally proceeded as though effective teachers can be identified and distinguished from ineffective teachers. Stodolsky (1984) challenged this assumption by arguing that teaching is a context-bound activity that varies considerably depending on the subject matter, instructional format, and objectives. Research examining the stability of teacher effects supports Stodolsky's argument. The question of whether a teacher who is effective in one situation is equally effective in other situations can be studied in three contexts: (1) when the same content is taught to different students either in different classes or across different years; (2) when different content is taught to the same students; (3) when different content is taught to different students.

Same Content, Different Students. Studies of teachers' effectiveness with the same content taught to different students have focused on long-term (periods of instruction stretching across several months) as well as short-term (periods of instruction lasting 30 minutes or less) effectiveness. Rosenshine (1970) reviewed four studies (Harris, et al., 1968; Morsh, Burgess & Smith, 1955; Soar, 1966; Torrance & Parent, 1966) that examined the long-term stability of teacher effects when the teacher taught the same material to different students, although none of the studies had focused on this topic as the major purpose of the research. Rosenshine concluded that these studies offered weak evidence for the stability of teacher effectiveness. Only the study by Harris et al. reported correlations as high as .5 and all other correlations were below .35. Rosenshine concluded that

> the lack of high stability coefficients in teacher effects may explain why studies of teacher characteristics have proven so futile. Teacher characteristics such as aptitude, attitudes, marital status, years of education, and number of courses in a given field are relatively stable. If these stable characteristics are correlated with unstable residual gain measures, we should expect `correlations that are nonsignificant, inconsistent from one study to the next, and usually lacking in psychological and educational meaning' [Gage, 1963, p. 118].

However this conclusion may be premature inasmuch as Rosenshine pointed out that these studies were subject to question due to the failure to assign students randomly to classrooms. Powerful uncontrolled variables such as student aptitude or socioeconomic status may have introduced systematic bias confounding the results. The use of standardized tests as the criterion of teacher effectiveness was an additional threat to the internal validity of these studies because these tests may not have measured the content covered in the teachers' instruction.

Rosenshine also reviewed a number of short-term studies conducted by Fortune (1966, 1967) in which the instructor taught 30 minutes or less. In five of the six samples that Fortune studied, the stability coefficients ranged from .45 to .70, with four of them significant at the .05 level. In striking contrast, in the long-term studies described above, only two of twelve correlations exceeded .40.

Research by Brophy (1973) suggests that individual teachers may differ in terms of the consistency of their effectiveness. He examined residual gain scores over 3 years for 165 elementary teachers. The effects of 28% of the teachers were consistent over the 3 years. The students of 14% of the teachers consistently achieved higher than expected in reading and mathematics; 14% consistently scored lower than predicted. The students' performance of 13% of the teachers improved consistently across the 3 years, while 11% consistently declined. Finally, students of the remaining 49% of the teachers performed inconsistently over the 3 years. A study by Emmer, Evertson, and Brophy (1979) offered further evidence that teachers vary in consistency and demonstrated support for Stodolsky's claim that stability varies with subject matter as well. The adjusted achievement of two classes taught by 39 English teachers and 29 mathematics teachers was compared. The students' California Achievement Test scores from the previous year were used to control for entering ability and knowledge. Achievement was measured by tests specially constructed to reflect the school district adopted curriculum. Intraclass correlations on the adjusted class means for the teachers' two classes were computed. Two coefficients were obtained: an estimate of the stability using a single class mean to estimate the teachers' effect and an estimate using the average of two classes' scores to estimate teachers' effects. For the 29 mathematics teachers, the correlations were .37 and .54 respectively, p<.021, and for the 39 English teachers, the coefficients were .05 and .10, p<.37. The stability of teacher effects increased markedly when teachers whose classes differed by 40 or more points were excluded from the sample. The values for mathematics teachers were .57 and .72, p<.002, and the values for English teachers were .29 and .45, p<.07. The strong correlation between the CAT and the students' achievement restricted the likelihood of finding high levels of stability. The correlation between the CAT and math achievement was .88 and the correlation between the CAT and English achievement was .94. Emmer et al. concluded that the stabilities in mathematics were high enough to warrant process-product research to identify variables related to student achievement (and the probability of finding reliable relationships could be increased by restricting differences in initial differences in ability between classes).

Different Content, Same Students. Fortune (1966a; 1966b; 1967) also examined the stability of teacher effects when different topics were taught to the same students. The correlations for the six studies described by Fortune and an additional study conducted by Belgard et al. (1968) on the same question ranged from -.27 to .47. The findings were surprising in that five of the fourteen correlations reported were negative, though insignificant. Berliner, Filby, Marliave, Moore, and Tikunoff (1976) studied 200 elementary school teachers who taught a 2-week unit in reading and mathematics. They found that the measures of effectiveness in the two subject areas correlated about .30.

Stodolsky (1984) also reported evidence that different content affects the stability of teacher effects. Trained observers recorded information about the activity structures of 20 fifth grade mathematics classes and 19 social studies classes. An average of 8.8 days of observations in the math classes and 8.1 days in social studies was obtained. Stodolsky concluded that subject matter was the major factor affecting variation in instruction. Mathematics instruction was relatively homogeneous within and across classrooms while social studies instruction was highly varied both within and across subject matter.

Different Content, Different Students. From the studies by Fortune and his associates correlations were also computed for the six samples of teachers when they taught different topics to different groups of students. Rosenshine concluded that these correlations were "the most perplexing of all" (p. 658), because they were unexpectedly higher in both directions than the correlations for teacher stability when teachers taught the same material to different students and when they taught different material to the same students. The correlations ranged from -.45 to .82. In contrast, Justiz (1960) found "amazing consistency" in two samples of student-teachers who taught two 30-minute lessons. The correlations were .63 and .90, both significant.

Conclusion. Rosenshine concluded that "the current long-term studies show that one cannot use the residual achievement gain scores in one year to predict the gain scores in a successive year with any confidence" (p.661). He recommended that stability estimates could be increased by using criterion measures that are more closely related to the content of instruction. The greatest degree of stability occurred in short-term situations in which the teacher instructed different groups of students on the same topic. In a more recent analysis of the stability of teacher effects, Berliner (1980) came to a similar conclusion that stability estimates are moderately stable when teachers teach the same content to similar students, but when different content is taught to similar students, the teacher effects do not appear to be stable.

Shavelson and Dempsey-Atwood (1976) concluded their review of the stability of measures of teaching behavior by stating that "generalizability may be extremely limited in an educational context" (p. 608). However, they qualified their conclusion by stating that the lack of stability may be due to the methodological inadequacies of the research rather than to the instability of teacher effectiveness. To shed light on this question, they conceptualized the issue of the stability of teacher effects in terms of generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) and recommended that studies of teacher effectiveness should vary systematically the situations across which policymakers intend to generalize. This would include classes, occasions, subject matter, and student abilites. Rowley's (1976) study of the generalizability of teachers' social orientation to students was cited as an example of how generalizability theory can be applied to determining the stability of teacher effectiveness.

In summary, the research suggests that teacher effectiveness may not be stable across different content and different students. Therefore, researchers should attempt to determine whether the relationships obtained between teacher education variables and student achievement are replicable across classes and subject matter. To obtain this evidence, longitudinal designs of educational effects are necessary.

Unit of Analysis

As noted in Chapter 2, many of the studies that have examined the relationship between teacher preparation and student achievement have used schools or districts rather than teachers as the unit of analysis. Veldman and Brophy (1974) pointed out that

> schools are not appropriate units for analysis [to show the effect that teachers have on student learning] because they are staffed by teachers of varying ability, and lumping together the data from these individual teachers masks rather than reveals the effects of the quality of schooling. Only data based on the teacher as the unit of analysis can show that some teachers are better than others. (p. 319)

Further, when the school is treated as the unit of analysis the impact of socioeconomic class is likely to be overestimated and the effect of the teacher underestimated because schools serving more economically advantaged students tend to have higher quality staffs (Burstein, 1980; Spady, 1976; Veldman & Brophy, 1974). Therefore, to obtain the best estimate of teacher effects, analyses should be conducted with the teacher as the unit of analysis for all of the input and output data.

A further problem related to the unit of analysis is the difficulty in interpreting results when the variables in the regression equation are aggregated at various levels of analysis. For example, when teacher-level variables are included in the same equation with school and district-level variables, it is difficult to interpret the results (Glasman & Biniaminov, 1981). To keep interpretation problems at a minimum, Cooley, Bond, and Mao (1981) recommended that the regression equation include variables at only one level higher than the dependent variable. Burstein (1980) described several approaches for analyzing data aggregated at more than one level. First, he described a model developed by Kiesling and Wiley (1974) to disentangle the effects of variables defined at one level from those defined at another level. Burstein also suggested using within-classroom slopes to deal with the problem. The third approach to this problem described by Burstein was developed by Kiesling (1978) and involves specifying different relational models for the between components and within components of the covariance matrix. This approach links the analysis of multilevel data to the developments in the analysis of covariance structures (Burstein, 1980).

In the analyses of teacher preparation effects on student achievement, the most appropriate unit of analysis is the teacher. Consequently, student data should be aggregated to the class level. However, in order to examine the possibility that school effects may mediate the relationship between teacher preparation and student achievement, it is necessary to include school-level effects in the analyses, as well.

The Problem of Multicollinearity in Regression Analysis

The validity of multiple regression analyses is jeopardized by the need to include highly correlated variables as predictors of student achievement. When two prediction variables are correlated in a multiple regression analysis, the first variable entered into the equation is likely to emerge as a significant predictor and when they are analyzed simultaneously only one variable tends to emerge as significant. For example, in Goodman's (1959) analysis when salary and education were entered simultaneously education emerged as the significant variable, but in other studies (Hanushek, 1972; Summers and Wolfe, 1975) salary emerged as the significant predictor. In addition, as Spady (1976) pointed out, under some statistical conditions, one variable may appear to have a positive relationship while the other seems to have a negative relationship. For example, Spady cited Armor's (1972) reanalysis of the Coleman et al. (1966) data, in which both teachers' salary and verbal ability tended to have a positive relationship to student achievement and teacher background, and school facilities were negatively related to achievement.

Multicollinearily is more likely to be a problem in aggregated data because in aggregating observations the random error component of the scores is likely to be cancelled (Asher, 1976, p. 48). The problem of obtaining spurious relationships as a result of collinearily can be reduced by increasing the size of the sample (Deegan, 1972). However, the inability of regression analyses to yield unequivocal results when input variables are correlated (an unavoidable condition in teacher effects research) demands that approaches to data analysis be identified that can avoid the multicollinearity problem.

Linear Analyses and Interaction Effects

Almost all studies of teacher effects have used analyses that examine only the additive relationships among variables. The assumption that there are no upper or lower limits to the relationships is unrealistic (Spady, 1976). Threshold effects are more likely. That is, increases in teacher variables like level of education or experience are likely to be related to student achievement up to a point beyond which further increases are likely to have no effect or perhaps even a negative effect

The consideration of possible interactive effects is also crucial and often overlooked in educational effects research.

Potter and Centra (1980) emphasized the importance of such analyses by citing the Summers and Wolfe (1975) study that found different effects for teacher experience at different levels of student achievment. That is, high-achieving students performed best with more experienced teachers while low achievers performed best with relatively inexperienced teachers. Spady (1976) found important interaction, threshold, accentuation, contextual, and curvilinear trends of this type in the existing school effects literature not reported in the original regression analyses by reanalysis of cross-tabular tables. Therefore, simple linear analyses are not adequate for the investigation of the complex relationships that exist among educational inputs and outputs.

Researchers should examine their data for complex effects. By carefully specifying relationships in the context of a theoretical model, the likelihood of identifying meaningful relationships will be increased.

The Shotgun Approach to Data Analysis

Previous input-output studies of educational effects have been characterized by a shotgun approach in which a large number of variables are included in the analyses in the hope that the analyses would reveal the relative importance of the variables. Pedhazur (1975) cautioned that "such a shotgun approach in a theoretical vacuum will not advance knowledge (p, 264). He emphasized that valid interpretations of school effects analyses require carefully specified equations that can be meaningfully interpreted within the framework of a substantive theoretical model.

Very little progress has been made in the development of a specific theory of schooling that can guide educational effects research. Biniaminov and Glasman (1983) identified only one study (Bidwell & Kasarda, 1975) that has tested a specific substantive model of school organizational effects. Levin (1980) described a conceptual framework that might be used to improve the estimation of educational production function. However, as Biniaminov and Glasman pointed out the complexities of "studying school variables at the secondary level" (p. 265) complicate the effort to design and test a theory of educational effects.

To guide research that can yield information useful in policy making, a model based on a theory of the relationship between teacher preparation variables and student achievement as it is mediated by school characteristics must be developed in concert with statistical procedures capable of analyzing educational effects at more than a single level.

A Methodological Synthesis

Although our literature review indicates that the research on teacher effects has been based on both the correlational and experimental paradigms of research described by Cronbach (1957), in the last 20 years the most popular approach to the study of educational effects has been the nonexperimental regression analyses adopted from econometrics, known popularly as input-output studies. Considerable controversy has surrounded the question of whether such analyses are appropriate for examining educational effects (Shapiro, 1984). Clearly, the difficulties of interpretation created by the problems of multicollinearity, the appropriate unit of analysis, sampling bias, and the adequate control of extraneous variables indicate that input-output studies alone will not improve our understanding of educational effects. Other factors that limit the usefulness of input-output studies are their inherent conservatism and, most important, their inability to reveal causal relationships.

The Conservatism of Input-Output Research

The correlational techniques used in input-output studies can only estimate relationships between variables as they are currently distributed in the schools. They are unable to estimate the potential effect if the values of the educational inputs were redistributed. For example, because of the current distribution of teachers holding a master's degree, it is unlikely that students will be assigned to master's level teachers consistently throughout the students' educational career. Therefore, it would be difficult to find students that would permit us to compare achievement of students who had been taught consistently by master's degree teachers with students taught exclusively by bachelor's level teachers. Thus, correlational techniques used in input-output studies are conservative strategies because they only permit examination of the status quo.

The Inability to Draw Causal Inferences from Input-Output Data

Some researchers have misused input-output analyses by drawing causal inferences from correlational data. For example, on the basis of input-output analyses of the Coleman Report, Levin (1970) suggested that "recruiting and retraining teachers with higher verbal scores is five to ten times as effective per dollar of teacher expenditure in raising achievement scores of students as the strategy of obtaining teachers with more experience" (p. 24). Such causal interpretations of correlational data are inappropriate and likely to lead to serious error, especially in cases like the Levin example, where implementation of his interpretation would have serious ramifications on hiring practices in education. The problem with such conclusions is that in nonexperimental research the relationship may be accounted for by a variable not included in the analysis. In the case of the Coleman Report for example, it appears likely that teachers with higher verbal ability were more often hired to teach in schools with high achieving students than teachers with lower verbal ability; therefore, increasing the number of teachers with high verbal ability may have no effect on student achievement. Policymakers can use regression techniques to draw conclusions regarding the investment necessary to produce specific effects in the dependent variables only when the data are derived from experimental designs (Pedhazur, 1975). In the case of the relationship between teachers' verbal ability and student achievement, only an experimental study in which the achievement of students randomly assigned to teachers with high verbal ablity is compared to the achievement of students randomly assigned to teachers with low verbal ability would warrant causal interpretations of the results.

The Power of Combining Experimental and Nonexperimental Designs

The strength of econometric methods is their ability to maximize external validity, because their goal is to identify relationships in a sample that can be generalized to the population (Shapiro, 1984). Consequently, econometric analyses can inform us about possible relationships and can be helpful in eliminating some rival hypotheses. However, input-output studies alone will leave us forever "founder[ing] in the swamp of uncontrolled plausible hypotheses" (Smith, 1972, p. 316). In contrast, experimental research can eliminate those plausible hypotheses. With experimental research, by virtue of the ability to manipulate the independent variables and control extraneous variables directly or by randomization, the researcher can draw causal inferences from the results, and when regression techniques are used to analyze data derived from experimental designs, policymakers can draw conclusions regarding the investment necessary to produce specific effects in the dependent variables.

Therefore, we believe the optimal approach to research of educational effects would unite the nonexperimental and experimental models and take advantage of the strengths of both approaches while compensating for the weaknesses of each. Although these two methodological approaches have never been used in concert within a single research design, an input-output study that examined various theoretical models could be followed by an experimental study that tested the causal direction of relationships obtained in the input-output phase of the research. Input-output research is more cost-effective during the exploratory phases of research, because it permits the investigation of a large number of variables including nonmanipulable ones. Such nonexperimental investigation is especially useful when variables are believed to have a causal effect on achievement, but some evidence of a relationship is needed before policymakers can be persuaded to increase the level of the variables as educational inputs. Experimental studies are labor- and cost-intensive and cannot feasibly be conducted to test all the relationships that researchers could conceive. Therefore, experimental designs should be reserved to test the relationships for which some correlational evidence exists to support the need for the study. Thus, research designs should make use of the unique strengths of each analytical approach. By testing alternative explanatory models with input-output analyses, potentially causal relationships can be identified that merit further investigation through more experimental procedures. Such an integration of research approaches is suggested in the design that we propose. First, we propose a traditional input-output study using causal modelling techniques to test a model of teacher preparation effects. Following the analysis of the data, we recommend the development of a study using a causal comparative design that further examines the relationships identified as important in the input-output study. For example, if level of education is found to be related to any of the educational outcomes, a study could be designed to compare the effectiveness of teachers holding a master's degree with teachers holding only a bachelor's degrees, controlling for variables like teacher experience, verbal ability, and socioeconomic status by selecting participants of equivalent experience, verbal ability, and socioeconomic status. Detailed observations and assessment of student performance could be made. Because various studies (Bassham, 1961, Perkes, 1967-68) have suggested that student ability may interact with teachers' educational preparation, student ability should be used as a blocking variable.

Summary and Implications

Much of the research conducted to date has been fraught with methodological weaknesses. The prevalence of these weaknesses among the studies reviewed limits the confidence that can be placed in these findings when drawing implications for policy or practice. The weaknesses noted in the existing body of research on effects of teacher preparation stem from three sources: (1) researchers used conveniently available data rather than collecting data in the form needed; (2) recently developed statistical procedures needed for appropriate data analyses were not widely available, when many of these studies were conducted; and (3) the scope of the study and sample were restricted because of inadequate resources. Major methodological problems can be summarized as follows:

> Sampling bias occurred in selection of teachers or inadequate numbers of teachers or schools were sampled to permit detection of effects at the classroom level.
>
> Teacher educational data were not collected or reported in sufficient detail to permit inferences that could guide future policies on teacher education.

Control for prior level of student achievement was inadequate.

Inadequate experimental or statistical controls for the effects of intervening variables (e.g., student SES, school characteristics, and teachers' level of motivation or sense of efficacy) that exert major influence on student achievement were incorporated into the studies.

Studies have been limited in scope, focusing only on one outcome measure or one grade level. Student attitudes have seldom been considered.

Student performance within a single study has been measured with different tests so that equating these scores is questionable.

Principal ratings (which are highly subjective) have often served as the outcome variable rather than objective measures of student outcomes.

Data were inappropriately analyzed using student score or school average as the unit of analysis. The most appropriate level of analysis, however, is the class average when inferences are to be drawn about effects of teacher characteristics.

The effects of correlated variables such as teacher ability, experience, teacher level of education, and teachers' salary frequently were confounded and the method of statistical analysis employed did not permit separate estimation of the effects of these variables.

> Previous input-output studies of educational effects included too many variables in regression analysis. Such a shotgun approach cannot significantly improve our understanding of how teacher education influences teachers' classroom effectiveness.

There is a clear need for a large-scale, comprehensive study of the relationship between critical variables in teacher preparation, school characteristics, and student performance. In Chapter 7 we describe how our proposal for the design of a research study of the relationship between teacher preparation and student achievement takes advantage of current state-of-the-art methodology to avoid the problems found in previous research in addressing the research question.

CHAPTER 7

A PROPOSED STUDY

The purpose of the proposed study is to develop and test a model for describing the relationship between teacher training factors, other teacher characteristics, school context variables, and student achievement. The study would use the teacher as the unit of analysis and the analysis of linear structural relationships to address the questions of interest. Because none of the teacher variables can be directly manipulated by the researchers the interpretations of relationships identified will be primarily statistical. An effort will be made, however, to control for irrelevant variation associated with student background and classroom/school context variables to strengthen the types of inferences which can be made about causal relationships between the variables.

The sections below present (1) an illustrative model specifying the types of variables to be studied and the hypothesized inter-relations; (2) examples of questions which will be answered; (3) methods and instruments for data collection; (4) a description of the sample and the minimum sample size needed; (5) a proposed data analysis strategy; and (6) a second phase of follow-up research, using a causal-comparative design, in which the most promising relationships identified in the model are subjected to more in-depth observation for a more restricted sample of teachers and students.

The Model

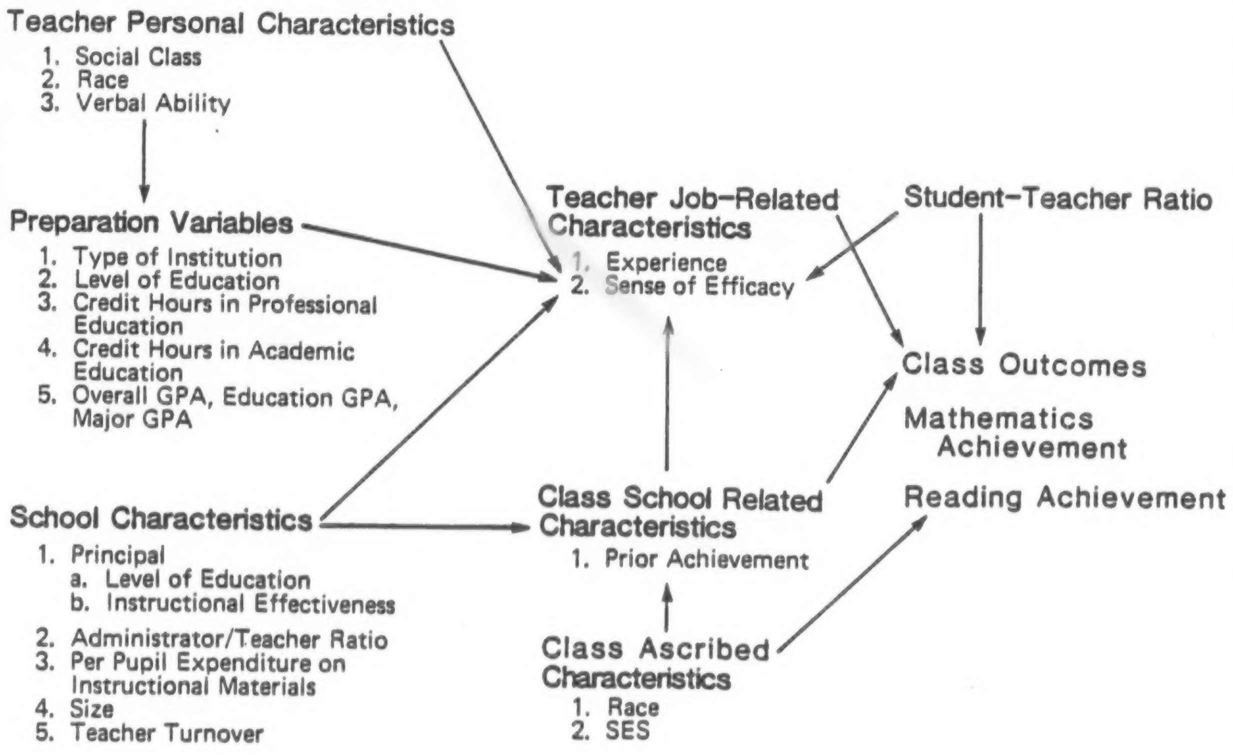
As pointed out in Chapter 6, Levin (1980) and others have emphasized that input-output studies failed to provide consistent and useful results, because they have not been based on a theoretical conception of educational effects. Researchers have relied solely on the empirical results of multiple regression analyses, and, consequently, have often obtained results that are difficult to interpret in the absence of a theory. To increase the likelihood that the proposed study will yield results that can guide the decisions of policymakers, we have developed a causal model of the relationship between teacher preparation variables and student achievement. The central organizing construct of the model is teachers' sense of efficacy. This construct has been shown in previous research to be significantly related to student achievement (Armor et al., 1966; Ashton & Webb, 1986; Berman et al., 1977; Gibson & Dembo, 1984, Glasman, 1984). Teachers' sense of efficacy refers to teachers' beliefs that they have the ability to teach and their students have the ability to learn. It has been hypothesized that teachers' efficacy beliefs affect student achievement because they influence teachers' "thoughts and feelings, their choice of activities, the amount of effort they expend, and the extent of their persistence in the fact of obstacles" (Ashton & Webb, 1986, p. 3). We expect that teacher preparation variables affect student achievement through the mediating influence of teachers' sense of efficacy. In other words, the experiences that teachers have in their teacher education programs create expectations in teachers regarding what they and their students are capable of accomplishing. These efficacy beliefs then influence teachers' classroom instruction and, ultimately, students' achievement. The model also reflects the effect that school and class characteristics can have on student achievement when moderated by teachers' sense of efficacy. For example, when principals reward and support their teachers for their performance, the teachers are likely to feel competent and appreciated and, therefore, increase their determination to teach effectively.

Figure 1 presents a diagram representing a theoretical model for explaining how teacher demographic characteristics, teacher education characteristics, school factors, sense of efficacy, and student characteristics combine to influence student achievement. Such a diagram is the first step to development of a structural equation model that can be used to assess the impact of these different variables on student achievement. In the language of structural equation modelling, an exogenous variable is an independent variable which is affected by no other variable in the model. In Figure 1, such variables have arrows flowing from them to other variables, but no arrow points toward an exogenous variable. An endogenous variable is a variable in the model which is affected by one of more other variables in the model. For example, teachers` level of education (an endogenous variable) may be affected by teacher verbal ability or teachers' SES (exogenous variables in the model). Student achievement is another endogenous variable that may be jointly affected by teachers' verbal ability and level of education. In Figure 1, endogenous variables have arrows pointing toward them. Note that it is possible for one endogenous variable to influence another in the model.

In formulation of a theoretical model, as a basis for a set of structural equations, an important issue is identification (Asher, 1976). The model depicted in Figure 1 leads to a set of equations which meet the criterion for identification. This is important because if a model is not identified, it is impossible to obtain a unique set of estimates of the parameters of that model.

Figure 1.

STRUCTURAL MODEL OF TEACHER PREPARATION VARIABLES INFLUENCING CLASSROOM OUTCOMES


Teacher Personal Characteristics
1. Social Class
2. Race
3. Verbal Ability
Preparation Variables
1. Type of Institution
2. Level of Education
3. Credit Hours in Professional Education
4. Credit Hours in Academic Education
5. Overall GPA, Education GPA, Major GPA
School Characteristics
1. Principal
a. Level of Education
b. Instructional Effectiveness
2. Administrator/Teacher Ratio
3. Per Pupil Expenditure on Instructional Materials
4. Size
5. Teacher Turnover
Teacher Job-Related Characteristics
1. Experience
2. Sense of Efficacy
Student-Teacher Ratio
Class Outcomes
Mathematics Achievement
Reading Achievement
Class School Related Characteristics
1. Prior Achievement
Class Ascribed Characteristics
1. Race
2. SES

As described earlier in our report, analyses conducted at different levels of aggregation address different questions. The most pertinent question for the investigation of teacher preparation effects must be investigated at the level of the teacher or class. However, we suggest an additional analysis at the level of schools. In other words, we recommend the investigation of a second structural model similar to that proposed in Figure 1 but conducted at the institutional level to explore the possibility that when the teacher preparation variables are aggregated to the school level the relationship may change from those that exist at the level of the teacher. For example, if we were to find no relationship between teachers' level of preparation at the class level, we might still find a relationship between these variables at the school level. This could occur if having a "critical mass" of master's level teachers in a school stimulates increased attention to student achievement and curriculum development, and this concern influences the instruction of bachelor's level teachers as well as master's level teachers.

The Questions

Several variations on the model depicted in Figure 1 would be developed, by systematically deleting some of the hypothesized relationships within the nested model so that the fit of the model to the data could be evaluated for successively simpler models. One question that could then be addressed through successive analysis is: Which of the several versions of the model fits the data best?

In addition to testing the overall fit of the data to the model, a series of questions following the paths depicted by the arrows in Figure 1 would be answered. For example, one series of questions would be:

> To what extent does teacher social class directly influence the level of education attained by the teacher?
>
> What is the direct effect of teachers' level of education on teachers' sense of efficacy?
>
> What is the direct effect of teachers' sense of efficacy on student achievement?
>
> What are the direct and indirect effects of teachers' level of education on student achievement?

Additional sets of questions would be answered for each possible path shown by the arrows that connect the variables in the model.

Instruments and Methods of Data Collection

As part of the present study we explored the feasibility of collecting accurate and timely information on teacher education and student achievement variables from data currently available through the State Department of Education. Data available from the Teacher Certification Office received particular scrutiny. We also consulted with school district personnel in several regions of the state to identify pragmatic procedures useful for collecting student achievement and teacher educational data at the district level. This information was taken into account in formulating the proposal.

From the feasibility study, we determined that the Comprehensive Test of Basic Skills (CTBS) is the most widely used standardized achievement test in the school districts in Florida. Student achievement test data can be collected from the county district office in the form of individual student test scores or the average test score for a given classroom. We also learned that the detailed information needed on teacher educational background cannot be obtained from existing data files of the Department of Education. Accurate, complete teacher educational background data can best be obtained from teachers directly. Furthermore, there is considerable variance in educational preparation of Florida's elementary teachers, but not in certification status of elementary teachers in Florida (Scott & Damico, 1985), so it is most reasonable to concentrate on differences in the type and amount of teachers' educational preparation.

The following instruments or methods of data collection would be employed:

1. A standardized achievement test with subscores in math and reading, such as the Comprehensive Test of Basic Skills, presently administered in 32 Florida counties;
2. A teacher questionnaire containing items relevant to the teacher's demographic and educational background;
3. A standardized measure of teachers' sense of efficacy (or motivation) such as that developed by Gibson & Dembo (1984).
4. A school questionnaire to be completed by the principal containing items on the school-level variables and classroom-level variables.

All items on these questionnaires would be pilot-tested for clarity of meaning and ease of response in at least two schools before being used in the field study. Teacher and school questionnaires would be distributed and collected by the research team on site in the schools. In addition, the Teacher

Questionnaire would contain a letter for the teacher's signature authorizing release of a transcript and admission test scores from the alma mater institution so that the educational background data (e.g., number of credits in professional education courses) could be obtained from a more accurate source than the teacher's memory.

Sample

To answer the research questions stated above, a random sample (or at least a representative sample) of teachers from at least two grade levels will be needed for the investigation. It is suggested that teachers who provide full time instruction at the second and fifth grades be included. While the choice of grade levels is arbitrary the selection of an early elementary and a late elementary grade level is recommended. The rationale for the choice is that (1) elementary grade level instruction is based on intact classrooms, (2) previous research has focused at these levels of instruction, (3) the grade level spread provides an opportunity to explore the generalizability of the results, (4) the inclusion of an early elementary grade reduced the confounding of multiple teacher effects. It might be possible to collect data on the teachers from the previous school year and explore the delayed effect of some teacher characteristics, and (5) the inclusion of the upper elementary grade will increase the variability in achievement test scores across classrooms.

It is estimated that approprimately 200 teachers at each grade level will be needed for the investigation. The 200 teachers at each grade level would be selected using a stratified sampling procedure so that approximately one-third should be from rural school districts, one-third from small metropolitan, and one-third from large metropolitan communities. A multistage sampling plan involving selection of districts within community-size strata and schools within districts would be used. Only the 32 districts which use the Comprehensive Test of Basic Skills would be included in the original population.

The minimum sample size depends on several factors including a) the minimum effect size to be detected which would be judged important from a practical point of view; b) the number of independent variables under investigation; c) the desired power level and d) the criterion for statistical significance. The use of these factors in determining the minimum sample size is explained below.

Effect Size: Effect size may be defined in terms of the proportion of variation explained by the predictor variables to the proportion of unexplained variation in the dependent measure. Cohen (1977) has provided some guidelines in defining effect sizes for multiple regression problems. Kraemer (1985) has also described effect size in terms of partial correlation coefficients. Specifying the minimal effect size which would be important to identify for this research is to a great extent arbitrary. We suggest that the sample size should be sufficient to permit detection of a partial correlation of .25 (or greater) between any of the independent variables and the outcome measure.

Independent Variables. The total number of independent variables can be divided into two groups. One group would consist of the control variables. These will include measures on student background (e.g., previous achievement, proportion of students receiving free lunches), classroom characteristics (e.g., number of students in the class, proportion of minority students) and school characteristics (e.g., total school population, present teacher turnover). The second group of independent variables includes teacher characteristics under investigation (e.g., possession of advanced degree, total graduate credits beyond the bachelor's degree, undergraduate major). The latter group of variables are the factors of primary interest in the investigation. In estimation of the required minimum sample size we arbitrarily designated that there would be approximately 5 teacher characteristics of primary interest. A moderate increase in this number would increase the necessary sample size only marginally.

Statistical Power. Statistical power is the probability that the null hypothesis will be rejected when it is in fact false. The null hypothesis which will be tested will state that teacher characteristics do not explain a significant proportion of variation in student achievement scores. If this hypothesis is in fact false we would like to be fairly confident that our analysis will result in rejecting it. Although there are few guidelines to define minimal power for studies such as the one proposed, it is recommended that a probability of .8 be accepted as a reasonable level of statistical power. Higher levels could be specified but the consequence would make the necessary sample size so large that the costs would be prohibitive. Lower power levels would be risky since important relationships between the variables might be missed.

Significance Level. In testing the null hypothesis for statistical significance the probability of Type I error will be set at the 5% level. Since the identification of false relationships between teacher characteristics and the student outcome could have detrimental effects on both teachers and students, this type of error should be minimized. The .05 level of significance is generally viewed as a reasonable criterion for testing statistical hypotheses.

Determining Sample Size. Taking the four factors into consideration Cohen (1977) presented a series of tables and formulas to estimate the minimal sample sizes for investigations using regression procedures. More recently, Kraemer (1985) has presented similar tables based on an approach which may yield more accurate estimates. Assuming that we desire that a partial correlation of .25 between one of the independent variables and an outcome variable should be statistically significant (for a power level of .80 and an alpha level of .05), Kraemer's table indicates that a minimum sample of 122 subjects is needed. Cohen's procedure yields a somewhat higher estimate (approximately 250 subjects). Both of these procedures must be considered as approximations for our model because structural equation coefficients are not, strictly speaking, quite the same as partial correlation coefficients. Nevertheless, these procedures provide some bases for estimating the minimum sample size that may be required in the proposed study. Based on these estimations, it seems that approximately 200 teachers at each grade level should be sampled, so that even if some attrition occurs, there would be an adequate number of teachers for the analysis. In arriving at these estimates it was necessary to assume that a simple random sample could be chosen from an infinite population. While this assumption may be violated in actual practice the above estimates should provide reasonable guidelines for determining the minimal number of teachers needed for the investigation.

Data Analysis

Once all the data have been collected calculations should be computed to describe the data distribution in terms of means and standard deviations for continuous variables and proportion of response frequencies for categorical variables. A correlation matrix should be developed and examined to eliminate or combine highly correlated variables.

The analysis would be conducted using LISREL VI, a program authored by Joreskog and Sorbom (1985), for the analysis of linear structural relationships. In simplified terms, this is a procedure for estimating the parameters of structural models and yields nonstandardized (or standardized) structural coefficients for various causal relationships hypothesized in the model of interest to the researcher. Specifically, the analysis would be used to (1) determine whether there is adequate fit between the data and the hypothesized model(s) and (2) test the significance of coefficients which quantify the degree of relationship between the outcome variables of interest (e.g., student achievement) and other variables in the model. The strength of this procedure lies in its ability to yield quantitative estimates of the direct and indirect relationships between variables within the specified model while taking into account how these variables are affected by other variables within the theoretical model that has been posited by the researcher. The limitation of this procedure is that interpretation of these coefficients rests upon the critical assumption that there is an adequate fit between the researcher's theoretical model and the empirical data. The proposed analysis would be replicated at each grade level.

Second Phase of Research

In the event that promising relationships are revealed in the analysis of the linear structural model, we recommend that a second phase of research explore the causal nature of the relationships through a causal-comparative research design similar to that developed by Popham (1971; see page 100 of this report) to measure teacher effectiveness. More intensive, detailed observations on this limited sample would permit the examination of teacher and student behavior as well as student achievement and student attitudes. Haertel (1986) has suggested a design which could be quite useful in this phase of the study. Further, we recommend that the stability of the effects be examined by including variations in terms of grade level, and subject matter replicated across time.

Time and Cost Estimates

The total time required to conduct a project such as that described, would be approximately 18 months. During the first

12-month period it would be reasonable to accomplish the major tasks of organizational start-up, questionnaire development, pilot-testing, drawing the sample, securing cooperation of participating districts and schools, collecting the teacher data, and obtaining student test-score data. The next 6 months would be devoted to data analysis and preparation of the final report.

The estimated cost for supporting the activities of the first 12 months would be approximately $85,000. The cost for supporting the major activities of the last 6 months of the 18-month project would be an additional $25,000. Thus the total cost of the 18-month project would be approximately $110,000. These cost estimates are based on the assumption that the work be conducted at one of the state universities or by an organization that would not charge for indirect costs.

References

Armor, D. (1972). School and family effects on black and white achievement: A reexamination of the USOE data. In F. Mosteller, & D.P. Moynihan (Eds.). On equality of educational opportunity (pp. 168-229). New York: Random House.

Armor, D., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L, Pascal, A., Pauly, E., & Zellman, G. (1967). Analysis of the School Preferred Reading Program in selected Los Angeles minority schools. (Report No. R-2007-LAUSD). Santa Monica, CA: The Rand Corporation. (ERIC Document Reproduction Service No. ED 130 243)

Asher, H. B. (1976). Causal modeling. Beverly Hills: Sage.

Ashton, P. T., & Webb, R. B. (1986). Making a difference: Teachers' sense of efficacy and student achievement. New York: Longman.

Bassham, H. C. (1960). Relationship of pupil gain in arithmetic achievement to certain teacher characteristics. Doctoral Dissertation, University of Nebraska.

Bassham, H. (1962). Teacher understanding and pupil efficiency in mathematics--a study of relationship. Arithmetic Teacher, 9, 383-387.

Beery, J. (1962). Does professional preparation make a difference? Journal of Teacher Education, 13(4), 386-396.

Begle, E. G. (1972). Teacher knowledge and student achievement in algebra (School Mathematics Study Group Reports No. 9). Stanford, CA: Stanford University.

Belgard, M., Rosenshine, B., & Gage, N. I. (1968). The teacher's effectiveness in explaining: Evidence on its generality and correlations with pupils' ratings and attention scores. In N. Gage et al. Explorations of the teacher's effectiveness in explaining. (ERIC Document Reproduction Service No. ED 028 147)

Benson, C. S. et al. (1965) State and local fiscal relationships in public education in California. Sacramento, CA: Senate of the State of California.

Berliner, D. (1980). Studying instruction in the elementary classroom. In R. Dreeben & J. Thomas (Eds.), The analysis of educational productivity. Vol. 1. Issues in microanalysis (pp. 191-222). Cambridge, MA: Ballinger.

Berliner, D., Filby, N., Marliave, R., & Weir, C. (1978). An intervention in classrooms using the Beginning Teacher Evaluation Study. Technical Report VI-1, Beginning Teacher Evaluation Study. San Francisco: Far West Laboratory for Educational Research and Development.

Burkhead, J., Fox, T. G., & Holland, J. W. (1967). Input and output in large city high schools. Syracuse, NY: Syracuse University Press.

Burstein, L. (1980). The roles of levels of analysis in the specification of educational effects. In R. Dreeben and J. A. Thomas (Eds.), The analysis of educational productivity. Vol.1: Issues in microanalysis, (pp. 119-190). Cambridge, MA: Ballinger.

Calabria, F. M. (1960). Characteristics of effective teachers. Educational Research Bulletin, 39, 92-100.

Carson, E. M. (1969). An analysis of National Teacher Examinations scores as predictors of teacher success in assignment. Unpublished doctoral dissertation, University of Houston.

Carter, J. C. (1967). Selected characteristics of beginning science and mathematics teachers in Georgia. Dissertation Abstracts International, 28(12), 4929A.

Caruthers, B. (1967). Teacher preparation and experience related to achievement of fifth grade pupils in science. Dissertation Abstracts International, 28(06), 1078A.

Carver, R. P. (1975). The Coleman Report: Using inappropriately designed achievement tests. American Educational Research Journal, 12, 77-86.

Centra, J. A. & Potter, D. A. (1980). School and teacher effects: An interrelational model. Review of Educational Research, 50, 273-292.

Clary, L. M. (1972). Teacher characteristics that predict successful reading instruction. Augusta College. (ERIC Document Reproduction Service No. ED 174 961)

Class, E. C. (1931). Prescription and election in elementary-school teacher training curricula in state teachers colleges. New York: Columbia University, Teachers College, Bureau of Publications.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Cohn, E., & Millman, S. D. (1975). Input-output analysis in public education. Cambridge, MA: Ballinger.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). Equality of educational opportunity. (2 Vols.) Office of Education, U.S. Department of Health, Education & Walfare, Washington, DC: U.S. Government Printing Office.

Cooley, W. W., Bond, L., & Mao, B. (1981). Analyzing multilevel data. In R. A. Berk (Ed.), Educational evaluation methodology: The state of the art. Baltimore: Johns Hopkins University Press.

Copley, P. O. (1974). A study of the effect of professional education courses in beginning teachers. Springfield, MO: Southwest Missouri State University. (ERIC Document Reproduction Service No. ED 098 147)

Cornett, L. M. (1984). A comparison of teacher certification test scores and performance evaluations for graduates in teacher education and in arts and sciences in three southern states. Atlanta: Southern Regional Education Board. (ERIC Document Reproduction Service No. ED 243 882)

Crane, E. (1958). Teachers' qualifications are improving. Bulletin to the Schools, 44, 173.

Cronbach, L. (1957). The two disciplines of scientific psychology. American Psychologist, 12, 671-684.

Cronbach, L., Gleser, G., Nanda, H., Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley.

Culpepper, J. G. (1972). A comparison of the academic preparation of high school biology teachers to student achievement in biology in selected south Arkansas school districts. Unpublished doctoral dissertation, University of Arkansas, Fayetteville.

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. Review of Educational Research, 53, 285-328.

Davis, C. R. (1964). Selected teaching-learning factors contributing to achievement in chemistry and physics. Unpublished doctoral dissertation, University of North Carolina at Chapel Hill.

Deegan, J. (1972). The effects of multicollinearity and specification error on models of political behavior. Unpublished doctoral dissertation. University of Michigan, Ann Arbor.

Delaney, E. C. (1954). Teacher selection and evaluation--with special attention to the validity of the personal interview and the National Teacher Examinations as used in one selected community (Elizabeth, New Jersey). Unpublished doctoral dissertation, Columbia University.

Denton, J. J., & Lacina, L. J. (1984). Quantity of professional education coursework linked with process measures of student teaching. Teacher Education and Practice, 39-46.

Druva, C. A. & Anderson, R. D. (1983). Science teacher characteristics by teacher behavior and student outcome: A meta-analysis of research. Journal of Research in Science Teaching, 20, 467-479.

Ducharme, R. J. (1970). Selected preservice factors related to success of the beginning teacher. Doctoral dissertation, Louisiana State Agricultural and Mechanical College.

Edelman, E. (1973). Levels of teacher reading-knowledge and pupil initial-status reading achievement: Their relation to levels of pupil residual-reading achievement gain. Dissertation Abstracts International, 35(02), 922A.

Eisenberg, T. A. (1977). Begle revisited: Teacher knowledge and student achievement in algebra. Journal for Research in Mathematics Education, 216-222.

Ellis, J. R. (1961). Relationships between aspects of preparation and measures of performance of secondary teachers of social studies. Journal of Educational Research, 55, 24-28.

Emmer, E.T., Evertson, C.M., & Brophy, J.E. (1979). Stability of teacher effects in junior high classrooms. American Educational Research Journal, 16, 71-75.

Flanagan, J. C. (1941). A preliminary study of the 1940 edition of the National Teacher Examinations. School and Society, 54, 59-64.

Fortune, J. C. (1966). A study of the generality of presenting behaviors in teaching preschool children. Memphis, TN: Memphis State University.

Fortune, J. C., Gage, N. I., & Shutes, R. E. (1966, February). The generality of the ability to explain. Paper presented to the American Educational Research Association. Amherst: University of Massachusetts, College of Education. (Mimeo.)

Frederickson, P. A. (1961). A study of teacher success measured by pupil achievement. Unpublished doctoral dissertation, Florida State University, Tallahassee.

General Accounting Office (1984, March). New directions for federal programs to aid mathematics and science teaching. Washington DC: General Accounting Office.

Gerlock, D. E. (1964). An analysis of administrators' evalu- of selected professionally certificated secondary school teachers. Doctoral dissertation, Florida State University.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. Journal of Educational Psychology, 76(4), 569-582.

Glasman, N. S. (1984). Student achievement and the school principal. Educational Evaluation and Policy Analysis, 6, 283-296.

Glasman, N. S., & Biniaminov, I. (1981). Input-output analyses of schools. Review of Educational Research, 51(4), 509-539.

Goldstein, H. (1985, April). Estimating the effects of schools in a multilevel variance components framework. Paper presented at the meeting of the American Educational Research Association, Chicago.

Goodman, S. M. (1959). The assessment of school quality. Albany, NY: State Department of Education.

Gray, H. B. (1962). A study of the outcomes of pre-service education associated with three levels of teacher certification. Doctoral dissertation, Florida State University.

Greenberg, J. D. (1983). The case for teacher education: Open and shut. Journal of Teacher Education, 34(4), 2-5.

Guthrie, J. W., et al. (1971). Schools and inequality. Cambridge, MA: Massachusetts Institutes of Technology Press.

Haertel, E. (1986). The valid use of student performance measures for teacher evaluation. Educational Evaluation and Policy Analysis, 8, 45-60.

Hall, H. O. (1962). Effectiveness of fully certified and provisionally certified first year teachers in teaching certain fundamental skills. Doctoral dissertation, University of Florida.

Hanushek, E. (1972). Education and race: An analysis of the educational production process. Lexington, MA: Lexington.

Harris, A. J., et al. (1968). A continuation of the CRAFT project: Comparing reading approaches with disadvantaged urban Negro children in primary grades. Cooperative Research Project No. 5-0570-2-12-1. Division of Teacher Education of the City University of New York. (ERIC Document Reproduction Service No. ED 010 297)

Hawk, P., Coble, C., & Swanson, M. (1985). Certification: It does matter. Journal of Teacher Education, 36(3), 13-15.

Hertz, W. S. (1959). The relationship between the teaching success of first-year elementary teachers and their undergraduate academic preparation. Dissertation Abstracts International, 20 (10), 4042.

Hice, J. E. L. (1970). The relationship between teacher characteristics and first-grade achievement. Dissertation Abstracts International, 31(08), 4036A.

Hurst, D. (1968). The relationship between certain teacher-related variables and student achievement in third grade arithmetic (Doctoral dissertation, Oklahoma State University, 1967). Dissertation Abstracts, 28, 4935A.

Joreskog, K. G., & Sorbom, D. (1985). LISREL VI user's guide. Mooresville, IN: Scientific Software.

Justiz, T. B. (1969). A reliable measure of teacher effectiveness. Educational Leadership, 3, 49-55.

Katzman, M. T. (1971). The political economy of urban schools. Cambridge, MA: Harvard University Press.

Keck, J. A. (1985). A historical review of the organization and development of teacher-education in the state of Florida: A case-study. Unpublished doctoral dissertation. University of Florida, Gainesville.

Kerlinger, F. (1973). Foundations of behavioral research. New York: Holt, Rinehart, & Winston.

Kiesling, H. J. (1969). The relationship of school input to public school performance in New York State. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Kiesling, H. J. (1970). The study of cost and quality of New York school districts: Final Report. Washington, DC: U.S. Department of Health, Education, & Welfare, Office of Education.

Kleyle, H. M. (1959). Differences in personal and professional characteristics of a selected group of elementary teachers with contrasting success records. Unpublished doctoral dissertation, University of Pittsburgh.

Kraemer, H. C. (1985). A strategy to teach the concept and application of power of statistical tests. Journal of Educational Statistics, 10, 173-195.

Lawrenz, F. (1975). The relationship between teacher characteristics and student achievement and attitude. Research Paper No. 8. (ERIC Document Reproduction Service No. ED 161 679)

Levin, H. M. (1970a). A cost-effectiveness analysis of teacher selection. Journal of Human Resources, 5, 24-33.

Levin, H. M. (1970b). A new model of school effectiveness. In A. Mood (Ed.), Do teachers make a difference? Washington, DC: U.S. Department of Health, Education and Welfare, Office of Education.

Levin, H. (1980). Educational production theory and teacher inputs. In C. Bidwell & D. Windham (Eds.), The analysis of educational productivity. Vol. 2. Issues in macroanalysis. (pp. 203-231). Cambridge, MA: Ballinger.

Lins, L. J. (1946). The prediction of teaching efficiency. Journal of Experimental Education, 15, 2-60.

LuPone, O. J. (1961). A comparison study of provisionally certified elementary school teachers in selected school districts in New York State. Journal of Educational Research, 55, 53-63.

Maguire, J. W. (1966). Factors in undergraduate teacher education related to success in teaching. Doctoral dissertation, Florida State University, 1966.

Massey, H. W. & Vineyard, E. E. (1958). Relationship between scholarship and first-year teaching success. Journal of Teacher Education, 9, 297-301.

Mayeske, G. W., et al. (1973). A study of the achievement of our nation's students. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Mayeske, G. W., & Beaton, A. E. (1975). Special studies of our nation's students. Washington, DC: U.S. Department of Health, Education, & Walfare, Office of Education.

McNeil, J. D. (1974). Who gets better results with young children--experienced teachers or novices? Elementary School Journal, 74, 447-451.

Medley, D. M., & Mitzel, H. E. (1963). Measuring classroom behavior by systematic observation. In N.L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally.

Michelson, S. (1970). The association of the teacher resourceness with children's characteristics. In A. Mood (Ed.), Do teachers make a difference? Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Moore, R. E. (1965). The mathematical understanding of the elementary school teacher as related to pupil achievement in intermediate-grade arithmetic. Unpublished doctoral dissertation, Stanford University.

Morsh, J. H., Burgess, G. C., & Smith, P. N. (1956). Student achievement as a measure of instructor effectiveness. Journal of Educational Psychology, 47, 79-88.

Murnane, R. J. (1975). The impact of school resources on the learning of inner city children. Cambridge, MA: Ballinger.

Nelson, B. J. (1978). The relationship of fifth- and sixth-grade students' achievement to pre-service science teacher preparation. Journal of Research in Science Teaching, 15(2), 161-166.

Norris, B. E. (1970). A study of the self-concept of secondary biology teachers and the relationship to student achievement and other teacher characteristics. Dissertation Abstracts International, 31(09), 4579A.

Norris, F. R. (1969). Pupil achievement as a function of an inservice training program on mathematics concepts for sixth grade teachers (Doctoral dissertation, George Peabody College for Teachers, 1968). Dissertation Abstracts International, 30, 1054A.

Osborn, C. E. (1970). A study of the qualifications of Mississippi high school biology teachers and the relationship of student achievement in biology to the subject matter preparation of the biology teacher. Dissertation Abstracts International, 31 (03), 1121A

Pedhazur, E. (1975). Analytic methods in studies of educational effects. In F. Kerlinger (Ed.), Review of research in education. Vol. 3 (pp. 243-266). Itasca, IL: Peacock.

Pedhazur, E. (1982). Multiple regression in behavioral research: Explanation and prediction. New York: Holt, Rinehart, & Winston.

Perkes, V. A. (1967-1968). Junior high school science teacher preparation, teaching behavior, and student achievement. Journal of Research in Science Teaching, 5(2), 121-126.

Perl, L. J. (1973). Family background, secondary school expenditure, student ability. Journal of Human Resources, 8, 156-180.

Peterson, D., Micceri, T., & Smith, B. O. (1985). Measurement of teacher performance: A study in instrument development. Teaching and Teacher Education, 1(1), 63-78.

Popham, W. J. (1971). Performance tests of teaching proficiency: Rationale, development, and validation. American Educational Research Journal, 8(1), 105-117.

Quirk, T. J., Witten, B. J., & Weenberg, S. F. (1973). Review of studies of the concurrent and predictive validity of the National Teacher Examinations. Review of Educational Research, 43, 89-113.

Romano, A. W. (1978). A study to determine the correlation between secondary teachers' biology knowledge and student achievement in biology. Dissertation Abstracts International, 39(10), 6047A.

Rosenshine, B. (1970). The stability of teacher effects upon student achievement. Review of Educational Research, 40(5), 647-662.

Rothman, A. I. (1969). Teacher characteristics and student learning. Journal of Research in Science Teaching, 6(4), 340-348.

Rothman, A., Welch, W., & Walberg, H. (1969). Physics teacher characteristics and student learning. Journal of Research in Science Teaching, 6, 59-63.

Rouse, W. M., Jr. (1968). A study of the correlation between the academic preparation of teachers of mathematics and the mathematics achievement of their students in kindergarten through grade eight. Unpublished doctoral dissertation, Michigan State University.

Rowley, G. (1976). The reliability of observational measures. American Educational Research Journal, 13(1), 51-59.

Ryans, D. G. (1951). The results of internal consistency and external validation procedures applied in the analysis of test items measuring professional information. Educational and Psychological Measurement, 11, 549-560.

Schlechty, P. C., & Vance, V. S. (1983). Recruitment, selection and retention: The shape of the teaching force. Elementary School Journal, 83(4), 469-487.

Schofield, H. (1981). Teacher effects on cognitive and affective pupil outcomes in elementary school mathematics. Journal of Educational Psychology, 73(4), 462-471.

Schunert, J. (1951). The association of mathematical achievement with certain factors resident in the teacher, in the teaching, in the pupil, and in the school. Journal of Experimental Education, 19, 219-238.

Scott, E., & Damico, S. (1985). The evaluation of the primary education act program and a comprehensive design for evaluation of the primary education act program (Final report, Contract No. 085-11). Department of Education.

Shapiro, J. (1984). On the application of econometric methodology to educational research: A meta-theoretical analysis. Educational Researcher, 13(2), 12-19.

Sharp, C. S. (1966). A study of certain teacher characteristics and behavior as factors affecting pupil achievement in high school biology. Dissertation Abstracts International, 1207A-1208A.

Shavelson, R., & Dempsey-Atwood, N. (1976). Generalizability of measures of teaching behavior. Review of Educational Research, 46(4), 553-611.

Shea, J. A. (1955). The predictive validity of various combinations of standardized tests and subjects for prognosis of teaching efficiency. Washington, DC: Catholic University of America Press.

Sheehan, D. S., & Marcus, M. (1978). Teacher performance on the National Teacher Examinations and student mathematics and vocabulary achievement. The Journal of Educational Research, 71, 134-136.

Shim, C. P. (1965). A study of four teacher characteristics on the achievement of elementary school pupils. Journal of Educational Research, 59, 33-34.

Shulman, L. S., & Carey, N. B. (1984). Psychology and the limitations of individual rationality: Implications for the study of reasoning and civility. Review of Educational Research, 54(4), 501-524.

Shuster, A. H., Jr. (1955). A study of certain aspects of a teacher education program at the Northern Illinois State Teachers College based upon a follow-up inquiry of beginning school teachers. Dissertation Abstracts International, 16, 691.

Shymansky, J. A., & Aldridge, B. (1982). The teacher crisis in secondary school science and mathematics. Educational Leadership, 39, 61-62.

Siegal, W. G. (1969). A study of the relationship between selected undergraduate academic schievement variables and teaching success. Doctoral dissertation, Washington State University.

Sirotnik, K. A. (1980). Psychometric implications of the unit-of-analysis "problem" (with examples from the measurement of organizational climate). Journal of Educational Measurement, 17, 245-281.

Smail, R. W. (1959). Relationships between pupil mean-gain in arithmetic and certain attributes of teachers. Unpublished doctoral dissertation, University of Denver, Denver.

Smith, M. (1972). Equality of educational opportunity: The basic findings reconsidered. On F. Mosteller & D. Moynihan (Eds.), On equality of educational opportunity (pp. 230-432). New York: Random House.

Soar, R. S. (1966). An integrative approach to classroom learning. Public Health Service Grant No. 5-R11 MH 01096 and National Institute of Mental Health Grant No. 7-R11-MH 02045. Temple University, Philadelphia, PA. (ERIC Document Reproduction Service No. ED 033 749)

Soeteber, W. H. (1969). Major-minor teaching assignments and related pupil achievement. Dissertation, Colorado State College.

Spady, W. G. (1976). The impact of school resources on students. In W.H. Sewell, R.M. Hauser, & D.L. Featherman (Eds.), Schooling and achievement in American society. New York: Academic Press.

Standlee, L. S., & Popham, W. J. (1958). Professional and academic preparation of teachers related to two indices of teaching performance (School of Education Research Report No. 3 [iv]). University of Indiana, School of Education.

Stodolsky, S. (1984). Teacher evaluation: The limits of looking. Educational Researcher, 13(9), 11-18.

Summers, A. A. & Wolfe, B. L. (1977). Do schools make a difference? American Economic Review, 67, 639-652.

Taylor, T. W. (1957). A study to determine the relationships between growth in interest and achievement of high school students and science teacher attitudes, preparation and experience. Doctoral dissertation, North Texas State College, Denton.

Thacker, J. A. (1965). A study of the relationship between principals' estimates of teaching efficiency scores on National Teachers Examinations, academic averages, and supervisors' estimates of potential for selected teachers in North Carolina. Dissertation Abstracts International, 26 (03), 1462.

Thoman, J. H. (1978). The relationships between teacher knowledge of science, preparation in science, teaching experience and fifth-grade achievement in science. Dissertation Abstracts International, 40 (05), 2578A.

Torrance, E. P., & Parent, E. (1966) Characteristics of mathematics teachers that effect students' learning. Cooperative Research Project No. 1020. Minnesota School Mathematics and Science Center, Institute of Technology, University of Minnesota. (ERIC Document Reproduction Service No. ED 010 378)

Tubbs, F. B. (1963). Some characteristics of highly effective and less effective secondary-school science teachers. Dissertation Abstracts International, 24(11), 4576.

Tuckman, H. P. (1971). High school inputs and their contribution to school performance. Journal of Human Resources, 6, 490-509.

Veenman, S. (1984). Perceived problems of beginning teachers. Review of Educational Research, 54, 143-178.

Veldman, D. J., & Brophy, J. E. (1974). Measuring teacher effects on pupil achievement. Journal of Educational Psychology, 66(3), 319-324.

Vockell, E. (1983). Educational research. New York: Macmillan.

Walberg, H. J. (1967). Scholastic aptitude, the National Teacher Examinations, and teaching success. Journal of Educational Research, 61, 129-131.

Walberg, H. J., & Rothman, A. I. (1969). Teacher achievement and student learning. Science Education, 53(3), 253-257.

Watts, G. E. (1964). A correlation analysis between "level of achievement" and certain teacher characteristics in selected school systems. Ohio University, Athens.

Wiley, D. E. (1976). Another hour, another day: Quantity of schooling, a potent path for policy. In W. H. Sewell, D. K. Featherman, & R. M. Hauser (Eds.), Schooling and achievement in American society. New York: Academic Press.

Willson, V. L., & Garibaldi, A. M. (1976). The association between teacher participation in NSF institutes and student achievement. Journal of Research in Science Teaching, 13(5), 431-439.

Winkler, D. (1972). The production of human capital: A study of minority achievement. Doctoral dissertation, University of California at Berkeley, 1972.

Winkler, D. R. (1975). Educational achievement and school peer group composition. Journal of Human Resources, 10, 189-205.

Table A-1

Summary of Studies of Level of Education and Teacher Effectiveness

Authors	Sample	Measure of Level of Education	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Bidwell & Kasarda (1975)	104 school districts in Colorado representing 90% of state student enrollment	% of teachers with master's degrees	Median %tile rank of secondary students in reading and math	District	Multiple regression	Sig. r for reading NS in math	1. Unit of analysis 2. No control for ability 3. Districts did not use same test
Brown & Saks (1975)	38 city, 116 suburban, & 365 town/rural districts in Michigan	% of teachers with master's degree	1. Average composite score in reading, math, and mechanics of written English 2. Test score variability	District	Multiple regression	Sig. r. for town/rural districts; NS for city & suburban; variability higher in suburb but not city and town/rural	1. Unit of analysis 2. Multicollinearity 3. Failure to control for initial ability
Burkhead, Fox, & Holland (1967)	39 high schools in Chicago	% of teachers with degrees beyond the bachelor's	1. IQ 2. reading	School	Stepwise regression	NS	1. Multicollinearity 2. Small sample size 3. Cross-sectional design 4. Multiple regression more appropriate analysis

Table A-1 (Cont'd.)

Authors	Sample	Measure of Level of Education	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Burkhead, Fox, & Holland (1967)	22 high schools in Atlanta	Median salaries	10th grade median verbal ability on SCAT	School	Stepwise regression	NS	1. Multi-collinearity 2. Small sample size 3. Cross-sectional design 4. Multiple regression more appropriate analysis
Burkhead, Fox, & Holland (1967)	177 small community high schools participating in Project Talent	Experience	Mean reading scores	School	Stepwise regression	NS	1. Multi-collinearity 2. Cross-sectional design 3. Multiple regression more appropriate analysis

Table A-1 (Cont'd.)

Authors	Sample	Measure of Level of Education	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Calabria (1960)	271 secondary teachers in New York State nominated as "effective" teachers	% of teachers with master's degree	Principal nomination of effective teachers	Teacher	Descriptive % of effective teachers	86% had Master's degrees compared to 33% of all teachers in state	1. No comparison group--A group of ineffective teachers might also have had a high % of Master's degrees 2. The definition of effectiveness was left to the principal to decide. 3. Self-selection; 21% of nominated teachers completed the questionnaire. 4. Did not consider student test performance.
Coleman, et al. (1966)	300 schools	% of teachers with degrees beyond the bachelor's	Verbal ability achievement test scores	District	Multiple regression	NS	1. Multicollinearity 2. Sample bias

Table A-1 (Cont'd.)

Authors	Sample	Measure of Level of Education	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Davis (1964)	18 secondary chemistry teachers-10 physics teachers	Bachelor's vs. master's degree	Standardized tests in chemistry and physics	Student	ANCOVA	1. Chemistry students achieved more when their teachers had Master's degree. 2. Physics students achieved more when teachers had bachelor's degree	1. Small sample 2. No control for teacher experience 3. Unit of experience analysis
Katzman (1971)	57 school districts in Boston	% of teachers with master's degrees	Achievement in mathematics and reading	School	Stepwise regression	NS in reading; neg r. in math	1. Multi-collinearity 2. Failure to control for initial ability 3. Unit of analysis 4. Small sample 5. Failure to control for student retention 6. Cross-sectional at single point in time 7. Difference scores in achievement

Table A-1 (Cont'd.)

Authors	Sample	Measure of Level of Education	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Kleyle (1959)	108 elementary teachers;	Credits earned beyond the bachelor's degree	Beecher Teaching Evaluations Rating	Teacher		NS	1. Student achievement test performance not considered 2. Questionable validity of ratings
Mollenhopt & Melville (1956)	100 schools 106 schools	% of teachers with 5 or more years of college teaching	7 types of mean achievement scores	District	Stepwise regression	NS	
Murnane (1975)	40 teachers/ 15 schools; 440 3rd graders; 442 students; studied in grades 2 and 3; middle class suburb of mid-west city	Master's degree	Reading, math, and spelling scores on Metropolitan Achievement Test	Student	Multiple regression	Inconclusive	1. Multicollinearity 2. Unit of analysis 3. Too many (18) independent variables 4. Few teachers with master's degrees 5. No control of previous grade teacher qualifications 6. No control for teacher attrition

Table A-1 (Cont'd.)

Authors	Sample	Measure of Level of Education	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Ober (1973)	58 elementary teachers from 11 elementary schools in middle class suburbs of a large mid-western city, 1449 students	Credits taken beyond bachelor's degree	Gain scores on Metropolitan reading and math tests	Teacher	Multiple regression	Significant relationship due to interaction between number of credits beyond BA/BS and years of teaching experience; NS on main effects	1. Small sample 2. Homogeneity of respondents
Perl (1973)	3,600 male senior high students (Project Talent sample); stratified random sample of 1000 high schools	% of teachers with master's degree; % with Ph.D.	Composite achievement scores: 1. Verbal ability 2. Abstract reasoning	District		M.A. related to abstract reasoning	1. Bias in sample undetermined 2. School as unit of analysis, ignores within school variations 3. Multicollinearity of certification and degree 4. No control for student entry-level ability
Peterson, Micceri, Smith (1985)	468 elementary teachers, 226 middle school teachers, 528 high school teachers	Bachelor's vs. Master's degree	Florida Performance Measurement System	Class	ANCOVA	NS	1. FPMS only appropriate for assessing first year teachers

Table A-1 (Cont'd.)

Authors	Sample	Measure of Level of Education	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Summers & Wolfe (1977)	627 6th grade pupils randomly selected from 103 randomly selected Philadelphia elementary schools	Education beyond the bachelor's degree	Composite achievement grade equivalent gain (ITBS)	School and student	Multiple regression	NS	1. Use of gain scores 2. Unit of analysis 3. Use of grade equivalents 4. Use of composite (math and reading) scores
Winkler (1975)	388 black and 385 white secondary students in a large urban school district in California In 1964-1965	Teacher salary	%tile rank of 8th grade reading achievement test	Student	Multiple regression	sig r. (reading achievement)	1. Multicollinearity of salary with experience and degree 2. Unit of analysis 3. Aggregated school level information 4. Characteristics only one achievement area

Table A-2

Summary of Studies of Teacher Preparation and Teacher Effectiveness

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Caruthers (1967)	5th grade teachers and their students	Number of hours in science courses	Math achievement	Class	ANCOVA t-test	Students whose teachers were experienced and had average of 18 hours in science courses had highest science achievement (inexperienced but prepared, second; experienced, and unprepared, third; inexperienced and unprepared, fourth	1. No control for extraneous variables
Copley (1974)	3 groups of beginning teachers, 22 liberal arts graduates with no professional education, 38 liberal arts graduates with education courses but no student teaching, 40 B.S. in Education graduates	Graduation from college of education or liberal arts (with and without student teaching)	Principal rating on 20-item scale	Teacher	Chi-square	Significant chi-square; education majors rated higher on interpersonal relations; NS on physical & mental health and personal qualities.	1. Subjectivity of rating scale

Table A-2 (Cont'd.)

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Culpepper (1972)	18 randomly selected teachers with 3-9 years experience from 30 southern Arkansas counties	Number of credit hours in biology (16 or fewer, 17-32, and 33-48 hours)	Scores on Nelson Biology Test of 20 randomly selected students from each teacher's class	Student	Multiple regression	Significant; Students whose teachers had more credit hours in biology had higher gain scores on on the NBT.	1.Small sample size
Davis (1964)	18 secondary chemistry and 10 secondary physics teachers	Graduation from liberal arts or teachers' college	Standardized tests in chemistry and physics	Student	ANCOVA	Higher achievement for students whose teachers graduated from liberal arts college	1.Small sample size 2.Not probability sample 3.No control for teacher experience
Davis (1964)	18 chemistry and 10 physics secondary school teachers	Number of credit hours in chemistry or physics courses	Standardized chemistry and and physics achievement tests	Student	ANCOVA	Significant negative relationship; students whose teachers had 10 or less hours in physics preparation and attended a NSF Institute scored higher on physics achievement test.	1.Small sample size 2.No control for teacher experience 3.Not probability sample

Table A-2 (Cont'd.)

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Denton & Lacina (1984)	55 education majors, 27 non-education majors;	Education or non-education major	1.Evaluation Profile-rating professional competence 2.Curriculum Content Checklist -rating planning effectiveness 3.Weekly Reflections Sheet (self-report, using time and morale)	Teacher	Sign tests	NS for planning or morale; Education majors rated higher on introducing and concluding lessons; non-education majors rated higher on use of duplicating and audiovisual equipment	1.Restriction in range of rating scale 2.Rating scale subjective
Ellis (1961)	70 experienced teachers	Graduation from college of education or liberal arts	Principals' nomination of "outstanding," "average," and "below average" teachers	Teacher	t-test	NS	1.Subjectivity of principal ranking 2.Identification of extreme groups
Ellis (1961)	70 experienced teachers	Amount of social studies preparation	Principal nomination of "outstanding", "average", and "below average" teachers	Teacher	t-test	NS	1.Subjectivity of principal ranking 2.Identification of extreme groups

Table A-2 (Cont'd.)

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Hice (1970)	40 first-grade teachers	Number of reading methous courses taken	Student mean adjusted end-of-year reading achievement scores on Metropolitan Readiness and Achievement Tests	Student	ANOVA	Sig. (Number of reading methods courses taken positively related to achievement of female students)	1.Use of categorical data 2.No control for student ability
Howe (1964)	51 10th-grade biology teachers randomly (stratified) selected from Oregon schools	Number of credit hours of preparation in science and biology	1.Knowledge and understanding of biological facts, concepts, and principles 2.Skills in applying methods of science 3.Improvement in critical thinking skills 4.Understanding of nature of science 5.Favorable attitudes toward science and scientific careers	Student	ANCOVA	Sig. (Students whose teachers had less than 40 hours in in science and less than 30 hours in biology did not rank in the upper third in gains in any of the 5 learning outcome criteria)	1.Use of categorical data
Hurst (1967)	55 3rd-grade teachers	Number of credit hours earned in mathematics education	Student mathematics achievement on Metropolitan Achievement Test	Student	ANOVA	NS	1.No control for student ability

Table A-2 (Cont'd.)

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Lawrenz (1975)	Stratified random sample of secondary science teachers - biology (84), chemistry (111), & physics (41). 60% response rate; no difference between respondents and nonrespondents	Number of credit hours in science methods	Scores of randomly selected class for each teacher on: 1. Learning Environment Inventory Test on Achievement in Science 2. Science Process Inventory 3. Science Attitude Inventory	Student	Stepwise regression	NS	1. Lack of fit between achievement test and student's science curriculum
Nelson (1978)	33 preservice teachers from 2 science methods courses randomly assigned to experimental (N=17) and control group (N=16) who were randomly assigned to groups of 5th and 6th grade students	45-minute instruction on teaching strategy for experimental group and no instruction for control group	Scores on 6-item test from Science: A Process Approach, Module 78	Student	ANOVA	Sig. (Students whose teachers received methods instruction had higher test scores than those whose teachers did not receive instruction)	1. Reliability of questionnaire questionable

Table A-2 (Contd.)

Authors	Sample	Criteria of Knowledge	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Lins (1946)	7 teachers from 5 subject areas	NTE	Average residual gain on standardized test	Student	Correlation	NS	1.Small sample size
Maguire (1966)	Elementary and secondary teachers	1.Overall GPA 2.GPA in general education 3.GPA in professional education 4.GPA in major 5.ACE scores 6.GPA in internship	Principal ratings	Teacher	Regression analysis	For first year teachers, significant results for internship, teaching field, and overall GPA. For 4th year secondary teachers negative results for general education GPA. No significance for elementary teachers	1.Subjectivity of principals' ratings
Massey & Vineyard (1958)	62 teachers	GPA	15 criteria of teaching effectiveness	Teacher	Correlation	Significant relationship of GPA with subject matter mastery, competence in English expression, general culture and character standards and ideals	1.High average GPA and ratings restricted the magnitude of the relationship

Table A-2 (Cont'd.)

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Osborn (1970)	High school biology teachers	Number of credit hours in biology (16 or fewer, 17-32, and 33-48 hours)	Scores on Nelson Biology Test	Student	Multiple regression	NS	1.Categorical measure of teacher preparation 2.Different groups of students used for pre- and post-tests
Perkes (1967-68)	32 junior high school teachers (population of junior high science teachers in a suburban California community)	Number of credits in science education (methods)	1.Scores of half the students on Sequential Test of Educational Progress Science Test Level 3 2.Scores of remaining half of the students on Junior High School Science Achievement Test	Student	Correlation	1.Sig (Number of credits earned in science education positively related to STEP test 2.Negative relationship between number of credits earned in science education and JHSSA (may have been stronger for students with middle to high IQ)	1.Small sample

Table A-2 (Cont'd.)

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Perkes (1967-68)	32 science teachers from 6 junior high schools in a suburban California community	Number of credit hours in science	Sequential Test of Educational Progress (Level 3)	Student	Correlation	NS	1.Small sample
Pisaro (1958)	196 unsatisfactory teachers and 168 superior teachers	Amount of college training and amount of professional education	71 Indiana school superintendents' descriptions of effective and ineffective teachers	Teacher	Descriptive	Teachers with less college training and less professional training more likely to be dismissed	1.Subjectivity of superintendent ratings
Rothman (1969)	51 physics teachers randomly selected from 17,000 physics teachers in the United States	1.Number of credit hours in physics 2.Number of credit hours in math	Physics Achievement Test, Classroom Climate, Welch Science Process Inventory, Attitude Questionnaire	Student	Canonical correlation	1.Significant relationship between number of credit hours in physics and student scores on PAT and interest in physics 2.Significant relationship between number of credit hours in math and student scores on TOUS, PAT, and interest in physics	

Table A-2 (Cont'd.)

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Rothman, Welch, & Walberg (1969)	35 male physics teachers (volunteers)	1.Number of credit hours in physics	1.Test of Understanding Science 2.Physics Achievement Test 3.Welch Science Process Inventory 4.Universe-beautiful and Physics-interesting subscales on Semantic Differential	Student	Canonical correlation	NS	1.Biased (volunteer) sample 2.Small sample size
Rouse (1967)	Mathematics teachers from kindergarten through grade 8	1. Number of credit hours in math	Math achievement	Student	Correlation	Low negative correlation between teachers' college math preparation and achievement from kindergarten to 6th grade and kindergarten to 8th grade	1.No control for student ability

Table A-2 (Cont'd.)

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Schunert (1951)	102 elementary algebra and 94 plane geometry teachers selected from 522 secondary schools listed in Minnesota Educational Directory	Graduation from state university, private college, or teacher's college	Achievement in geometry and algebra (used researcher-developed tests)	Student	ANCOVA	Algebra achievement of graduates of state universities and private colleges higher than graduates of teacher colleges; geometry achievement-NS	1.No control for teacher ability
Sharp (1966)	High school biology teachers	Number of semester hours in biology, chemistry, and physics	Scores on Nelson Biology Test	Student	Multiple regression	NS	1.Lack of control for student ability
Smail (1959)	97 4th, 5th, & 6th grade teachers in Souix Falls, SD public schools	Number of years of college education (2 vs. 4)	Mean gain in in arithmetic	Student		NS	1.Use of gain scores
Soeteber (1969)	34 algebra teachers randomly selected from 15 Wisconsin school systems	Number of credit hours in math (37 hours or more vs. 36 hours or less)	Scores on Algebra I test	Student	ANOVA	NS	1.Inappropriate unit of analysis 2.Small sample 3.Restriction in range of Algebra I test scores

Table A-2 (Cont'd.)

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Standlee & Popham (1958)	880 teachers who graduated in 1965 from 24 Indiana colleges and universities with teacher education accreditation	Graduation from small public & private and large public and private institutions	Teacher ranking form (principals rank order teachers)	Teacher	% judged by principals to be higher in overall teacher effectiveness	Higher proportion graduated from small public or large private institutions than from public or small private ones	1.Subjectivity of principal ranking
Standlee & Popham (1958)	(See above)	Number of credit hours in professional courses	Principal rankings	Teacher	Chi-square	NS	1.Subjectivity of principal rankings
Summers & Wolfe (1975)	627 6th-grade pupils randomly selected from 103 randomly selected Philadelphia elementary schools	"Quality" rating of teachers' undergraduate college	Composite achievement grade equivalent gain ITBS	Student	Multiple regression	Positive relationship between student gains and "quality" rating of teacher's college	1.Use of gain scores 2.Use of grade-equivalents 3.Use of composite scores
				School	Multiple regression	NS	Same as above

Table A-2 (Cont'd.)

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Taylor (1957)	83 science teachers (42 fulltime, 41 parttime science teachers in grades 9-12)	1.Number of semester hours of professional education 2.Number of credit hours in science	1.Student achievement score on Essential High School Content Battery 2.Student interest-Occupational Interest Inventory	Student	Correlation	1.Zero order NS; science achievement of students whose teachers were above median on 3 or 4 factors (hours of professional education and science course, years of experience, MTAI) significantly higher than achievement of students whose teachers fell below the median 2.Students whose teachers were above the median number of credit hours in science courses (45.5) scored significantly higher on the achievement test than those whose teachers were below the median ($p < .10$)	1.Identification of extreme groups 2.Use of one test to assess achievement in general science, biology, chemistry, and physics 3.No control for student ability

Table A-2 (Cont'd.)

Authors	Sample	Criteria of Preparation	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Thoman (1978)	29 teachers randomly selected from 5th grade classes in southeastern Wisconsin	Number of credit hours in science	Achievement gain on STEP Science Test	Class	Multiple correlation	NS	1.Small sample 2.No control for student ability 3.Possible range restriction (of credit hours in science)
Winkler (1975)	388 black and 385 white secondary students in a large urban California school distrist in 1964-1965	Graduation from prestigious university	% rank of 8th grade reading achievement test	Student	Multiple regression	Positive relationship between achievement and graduation from a prestigious university	1.Multi-collinearity 2.Only one achievement area 3.Aggregated school level information

Table A-3

Summary of Studies of Teacher Knowledge and Teacher Effectiveness

Authors	Sample	Criteria of Knowledge	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Bassham (1961)	28 6th grade teachers, 620 students	Test of Basic Mathematical Understanding	California Achievement Test, Mathematics (Form AA in Sept., Form BB in April); Henmon-Nelson Test of Mental Ability (Fall); Arithmetic Interest Inventory (Sept. and April)	Student	Multiple correlation	Significant positive relationship for students with above average intelligence	1.Small sample size
Begle (1972)	308 volunteer 9th grade teachers--NSF institute participants	Algebraic Inventory	Mathematics Inventory (Fall); Reference Test for Cognitive Factors (Fall); Computation and noncomputation tests (Spring)	Student	Stepwise Regression	NS	1.Bias of sample
Clary (1972)	23 4th grade teachers in Spartanburg, S.C.	Inventory of Teacher Knowledge of Reading	Science Research Associates Achievement Series, Reading (alternate forms in both Oct. and March)	Student	Stepwise Regression	Significant (the best predictors of reading achievement are personality and knowledge of reading)	1.Small sample size 2.Unvalidated measure of knowledge

Table A-3 (Cont'd.)

Authors	Sample	Criteria of Knowledge	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Ducharme, Sheehan, & Marcus (1978)	119 first-grade teachers (1836 pupils)	NTE-WCET	Metropolitan Readiness, Word Meaning & Number subtests; Iowa Test of Basic Skills	Class	Stepwise Regression	NS	1.Multi-collinearity
Edelman (1973)	200 teachers of grades 4-8 in Chicago public schools	Inventory of Teacher Knowledge of Reading	Metropolitan Reading and Word Knowledge Subtests	Student	Partial correlation; chi-square	No r for continuous data; categorical analysis showed interaction between teacher knowledge and students' initial achievement status	

Table A-3 (Cont'd.)

Authors	Sample	Criteria of Knowledge	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Eisenberg (1977)	28 Algebra 1 teachers in Columbus, Ohio	Algebraic Inventory	Mathematics Inventory	Student	Multiple Regression	NS	1.Bias of sample--10 principals refused to allow teachers to participate and 9 refused to participate 2.Small sample size (28 subjects, 15 variables)
Hertz (1959)	157 first year elementary teachers	GPA	Principal rating	Teacher	Correlation	Teachers in top 40% of class received higher ratings	1.Lack of reliability of rating scale
Kleyle (1959)	108 elementary teachers	1. NTE 2. GPA	Beecher Teaching Evaluation Record	Teacher	Multiple correlation	Significant relationship for NTE; NS for GPA	1.Subjectivity of ratings
Lawrenz (1975)	Stratified random sample of secondary biology (84), chemistry (111) and physics (41) teachers	NTE in Science; Science Process Inventory; Science Attitude Inventory	Tests of achievement in science	Student	Stepwise regression	Significant positive relationship between teachers' scores on Science Process Inventory and student achievement	1.Multicollinearity

Table A-3 (Cont'd.)

Authors	Sample	Criteria of Knowledge	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Norris (1968)	18 6th grade mathematics teachers who attended inservice compared with 15 6th grade teachers who did not and 702 pupils randomly selected from their classes.	Participation in inservice	Specially constructed 50-item multiple-choice concept test	Student	ANCOVA	Pupils of inservice group scored significantly higher; no difference between teacher group scores on posttest (60-item multiple-choice test (though mean gains of experimental teachers were greater)	1.Administration of teacher unsupervised. 2.Students atypical; top 22% of students in prior math achievement; 3.Teachers volunteered for participation in inservice; 4.No control for initial differences in teacher knowledge (experimental teachers had lower initial scores on pretest).
Norris (1970)	30 teachers attending NSF institutes at Ball State University	1.Tennessee Self-Concept 2.Commission on Undergraduate Education in Biological Sciences Tests	1.Processes of Science Test 2.Differential Aptitude Test	Class	Multiple correlation	Significant positive relationship between teacher scores on biology knowledge test and teaching proficiency scores	1.Sample bias

Table A-3 (Cont'd.)

Authors	Sample	Criteria of Knowledge	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Perkes (1967-68)	32 teachers from a population of science teachers from 6 junior high schools in a suburban California	GPA	1.STEP: Science Test Level 3 2.JHSSA	Student	Correlation	Significant positive relationship between teacher GPA and STEP; significant relationship between teacher GPA and JHSSA (may be stronger for higher IQ courses)	1.No control for student ability 2.Multicollinearily between GPA and teaching style 3.GPA confounded with number of credits in science courses
Romano (1978)	50 secondary teachers randomly selected from a group of 257 who returned their NTE scores; 35 with complete data	NTE (Biology)	Residual gain scores on the Cooperative Science Test--Biology Forms A and B	Student	Correlation	Positive relationship	1.Inconsistency in degree to which teachers taught material covered on criterion test 2.No control for student ability 3.Sampling bias

Table A-3 (Cont'd.)

Authors	Sample	Criteria of Knowledge	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Rothman (1969)	51 teachers selected randomly from a national pool of 17,000 physics teachers in the U.S.	1.Teacher knowledge & student TOUS scores 2.Test on selected topics in physics 3.Test on Understanding Science 4.Number of semester hours in physics & physics education 5.Number of hours in math	Physics Achievement Test, Classroom Climate Questionnaire, Welch Science Process Inventory, Attitude Questionnaire	Student	Canonical correlation	Significant positive relationship between teacher knowledge & student scores on Test of Understanding Science	
Rothman, Welch,& Walberg (1969)	35 male physics teachers (volunteers)	Scores on test on Selected Topics in Physics	1.Test of Understanding Science 2.Physics Achievement Test 3.Welch Science Process Inventory 4.Tinkering Subscale of Pupil Activity Inventory 5.Universe-beautiful and Physics-interesting Semantic Differential Subscales	Student	Canonical correlation	NS	1.Small sample size 2.Biased sample

Table A-3 (Cont'd.)

Authors	Sample	Criteria of Knowledge	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Sharp (1966)	High school biology teachers	NTE	Nelson Biology Test	Student	Multiple regression	NS	1.No control for student ability
Siegel (1969)	Phase 1:1,010 first-year teachers (393 elementary & 617 secondary teachers)	1.Undergraduate GPA 2.Professional education (Ed.GPA) 3.Non-education GPA 4.Major field GPA	Principal ratings on Beecher's Teaching Evaluation Record	Teacher	Correlation	For elementary teachers, significant relationship between GPA (.10) and Ed.GPA (.11). For secondary teachers, significant relationship for UG-GPA (.20), Ed-GPA (.16), Non-ed-GPA (.13), and major field GPA (.18).	1.Subjectivity of ratings
Siegel (1969)	Phase 2: Elementary and secondary teachers	1.GPA 2.Secondary area major coursework 3.Methods courses	Principal ratings	Teacher	Partial correlations holding undergraduate GPA constant	NS	1.Subjectivity of ratings

Table A-3 (Cont'd.)

Authors	Sample	Criteria of Knowledge	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Smail (1959)	97 teachers of grades 4-6 in Sioux Falls, SD	Test of knowledge of basic math concepts	Mean gain in arithmetic	Student	Multiple regression	NS	1.Use of gain scores
Soeteber (1969)	22 Algebra I teachers (1,184 students)	Standardized Algebra test--Advanced Algebra III	Algebra I test specially designed for this study	Student	ANOVA	Significant relationship between teacher scores on advanced algebra test and student achievement	1.Inappropriate unit of analysis 2.Small sample 3.Restriction in range of algebra I test scores
Thoman (1978)	29 5th grade teachers randomly selected from the population of 5th grade teachers in Southeastern Wisconsin	STEP 1A	STEP Series II gain scores	Class	Multiple correlation	NS	1.No control for student ability 2.Data on teachers' scores not provided to determine if ceiling effect restricted magnitude of relationship 3.Multicollinearity of teacher knowledge with the predictors

Table A-3 (Cont'd.)

Authors	Sample	Criteria of Knowledge	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Walberg & Rothman (1969)	36 teachers selected from 500 who volunteered to field test the Harvard Project Physics, a new physics course	36 items from the unit tests of the Harvard Project Physics	Physics Achievement Test, Classroom Climate Questionnaire, Welch Science Process Inventory Attitude Questionnaire	Student	Multiple regression	Significant negative relationship: teachers with higher achievement gave lower grades	1. Sample not representative of science teachers 2. Order of variables in regression model may have affected results 3. Adequacy of measure of teachers' achievement in question

Table A-4

Summary of Studies of Certification and Teacher Effectiveness

Authors	Sample	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Beery (1962)	76 1st year provisionally certified teachers (none had student teaching, 34 had no professional education courses, 42 had at least one course), 76 matched teachers with regular certification	5 observations by professionals using the Ryans' Classroom Observation Record	Class	t-test	Of 45 comparisons, fully certified teachers superior on all, 25 of comparisons were significant	1.Matching on age and years since graduation inadequate, fully certified teachers were older
Bledsoe, Cox & Burnham (1967)	Randomly selected provisionally and certified teachers in science, social studies, English and mathematics at secondary level, elementary teachers of grades 1-6 in Georgia	33 self-report and classroom behavior variables	Teacher	t-test	Fully certified teachers rated higher on 11 of 33 criteria; more systematic and responsible, more skilled in the use of teaching media, more competent in non-specific teaching behavior, more generally competent, more satisfied with teaching and their profession	1.Subjectivity of ratings and self-report

Table A-4 (Cont'd.)

Authors	Sample	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknessess
Cornett (1984)	Provisionally certified teachers and fully certified teachers in Georgia in 1982-83	Georgia Teacher Certification Tests	Not applicable	No analysis	Teachers with Bachelor's degree scored lower than Master's level teachers chers except in science (no test of significance of differences)	1. Restriction in range of of scores 2. Small number of teachers
Cornett (1984)	All provisionally certified teachers in Louisiana in 1982-83 (N=89) 105 regularly certified teachers selected by random sample	National Teacher Examinations 1.Weighted Common Examinations 2.Area Exams	Not applicable	No analysis	1.Teachers with temporary certificates scored higher on the WCET (619 compared to 602) 2.On Elementary Area Exam, regularly certified scored higher than temporarily certified	1.Temporarily certified had completed from 0 to 36 credit hours in education

Table A-4 (Cont'd.)

Authors	Sample	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Cornett (1984)	All teachers (N=21) holding provisional certificates in Georgia during 1982-83 and a matched group of regularly certified teachers	Locally developed teacher evaluation instrument for assessing beginning teachers on 10 competencies	Not applicable	No analysis	1.Provisionally certified scored 150 out possible 165; regularly certified scored 158	1.Fully certified averaged 3 years more experience than provisionally certified teachers 2.No statistical test of differences 3.Ceiling effect on evaluation instrument
Cornett (1984)	All teachers holding provisional certificates from 1979-83 (N=191) in LA. and random sample of 348 certified teachers	Principal rating on 33 basic teaching functions	Not applicable	No analysis	No differences	1.Lack of variance in ratings 2.Subjectivity of ratings
Gerlock (1964)	341 beginning secondary teachers--201 professionally certified--140 provisionally certified	Administrator ratings of: 1.Personal qualifications 2.Teaching skills 3.Relationships with others 4.Professional ethics 5.Moral and social ethics	Teacher	Chi-Square	Professionally certified rated higher in teaching skills ability, moral, and social ethics, and observing	1.Use of unvalidated rating scale

Table A-4 (Cont'd.)

Authors	Sample	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Gray (1962)	763 beginning white teachers in Florida--110 holding a temporary certificate --100 holding a regular certificate	1. teacher self-evaluation 2. principal evaluation 3. MTAI	Teacher	Chi-square	Significant on all 3 measures of effectiveness	1. Subjectivity of ratings
Hall (1962)	38 elementary teachers--21 provisionally certified and 17 fully certified	Grade equivalent gain scores on the Stanford Achievement Test: paragraph meaning, word meaning, spelling, language, arithmetic reasoning, and arithematic computation	Class	ANOVA	Pupils of certified teachers gained significantly more in spelling, with similar trends for paragraph meaning and word meaning	1. Use of gain scores 2. Small sample
Hawk, Coble, & Swanson (1985)	36 middle and high school teachers in grades 6-12. 18 in-field and 18 out-of-field pairs teaching the same subject to students of same ability level at at same school	1.Student achievement--Stanford Achievement Test, General Math, & Stanford Test of Academic Skills (algebra) 2.Teacher knowledge-Descriptive Tests of Mathematics Skills 3.Professional Teaching Skills--2 observations during a 7-month period, using Carolina Teacher Performance Assessment System (CTPAS)	Student	ANOVA	1.Achievement was higher in general math and algebra classes taught by certified teachers 2. In-field teachers scored higher in algebra achievement 3. Certified teachers had higher scores on instructional presentation	1. Small sample

Table A-4 (Cont'd.)

Authors	Sample	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Lupone (1961)	From a sample of 240 teachers in their 1st, 2nd, and 3rd years of teaching in selected school districts in New York state, 40 provisionally certified and 40 fully certified teachers were compared.	Principal rating	Teacher	t-test	Permanently certified were were more effective in 5 of 7 areas rated: 1.Preparation; 2.Planning and management; 3.Subject matter 4.Pupil-teacher relations 5.Evaluation. During 2nd and 3rd year permanently certified rated superior to provisionally certified in instruction. No differences in human relations.	1.Subjectivity of principal ratings

Table A-4 (Cont'd.)

Authors	Sample	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
McNeil (1974)	19 experienced teachers of grades K-6, 19 beginning education students	1. Student success on a criterion-referenced test 2. Student rating of interest	Class	Chi-square	1. Students of experienced teachers scored higher on the test & expressed greater interest	1. No control for prior ability 2. Use of categorical data
Massey & Vineyard (1958)	62 teachers	Supervisor ratings	Teacher	Correlation	Regularly certified teachers received higher ratings than provisionally certified teachers	1. High average rating restricted the likelihood of finding differences
Perl (1973)	3600 male senior high students	1. verbal ability 2. abstract reasoning	High School	Multiple Regression	No relationship	1. Bias in sample undetermined 2. Within school variations ignored

Table A-4 (Cont'd.)

Authors	Sample	Criteria of Effectiveness	Unit of Analysis	Analysis	Results	Weaknesses
Popham (1971)	28 pairs of credentialed and non-credentialed teachers in auto mechanics, 16 in electronics, 13 in social studies	Criterion-referenced achievement test	Class	ANOVA	No difference	1.Questionable validity of criterion tests 2.Certified teachers were in classroom while uncertified teachers taught 3.Multicollinearity of certification, degree, % of time spent in area of specialization, experience
Shim (1965)	89 teachers from semi-rural district	California Achievement Tests, Form W	Class	t-test	No difference, although fully certified teachers taught less able students	1.Failure to control for ability level of students 2.Use of categorical data 3.Multiple t-tests